

COMSAT Laboratories

Technical Report No. CL-TR-5-71

ORTHOGONAL TRANSFORM
FEASIBILITY STUDY

November 1971

Final Report
Prepared for
NASA
Manned Spacecraft Center
Under Contract NAS 9-11240



W. L. Pritchard
Ass't Vice President and
Director of COMSAT Labs

The work described in this Final Report was supported by NASA under Contract NAS 9-11240 and by COMSAT Laboratories under the on-going research program at the Laboratories.

Principal Investigator:
Dr. Guner S. Robinson

Project Manager:
Dr. S. J. Campanella

Table of Contents

	<u>Page No.</u>
1. INTRODUCTION	1-1
2. RATE-DISTORTION CRITERIA	2-1
2.1 General Rate-Distortion Criteria	2-1
2.2 Rate Distortion for a Discrete Memoryless Source	2-2
2.3 Application to Speech and Visual Signal Processing	2-4
3. ORTHOGONAL TRANSFORMATION TECHNIQUES FOR ONE DIMENSION	3-1
3.1 General	3-1
3.2 The Discrete Fourier Transform (DFT)	3-1
3.3 The Discrete Walsh-Hadamard Transform (DWT) ..	3-7
3.3.1 Hadamard Matrices	3-7
3.3.2 A Computational Algorithm for Hadamard Transform	3-16
3.4 The Discrete Haar Transform (DHT)	3-22
3.4.1 Haar Functions	3-22
3.4.2 Haar Transform Algorithm	3-25
3.4.3 Inverse Haar Transformation	3-26
3.5 The Discrete Karhunen-Loève Transformation-- The Optimum Orthogonal Transformation	3-28
3.5.1 General	3-28
3.5.2 Karhunen-Loève Transform Development ..	3-28
4. ORTHOGONAL TRANSFORM PROCESSING OF SPEECH	4-1
4.1 General	4-1
4.2 Information Rate Reduction Using Orthogonal Transforms	4-1

Table of Contents (Continued)

	<u>Page No.</u>
4.3 Speech Sampling and Analysis of Window Size ..	4-6
4.4 Fourier Transform Processing of Speech	4-8
4.5 Hadamard Transform Processing of Speech	4-13
4.6 Haar Transform Processing of Speech	4-14
4.7 Karhunen-Loève Transform Processing of Speech	4-17
4.7.1 Introduction	4-17
4.7.2 Determination of the KLT Matrix	4-18
4.7.3 Results of Speech Processing Using the KLT	4-22
4.8 PB Word List Processing and Testing	4-24
4.9 Additional Attempts to Reduce Bit Rate	4-31
4.9.1 Variation of Coefficient Bit Assignments	4-32
4.9.2 Polar Coordinate Representation	4-36
4.10 Summary of the Orthogonal Transform Tech- niques for Speech Processing	4-39
5. ORTHOGONAL TRANSFORMATION TECHNIQUES FOR TWO DIMENSIONS	5-1
5.1 Introduction	5-1
5.2 PCM Image Processing	5-3
5.3 Quantizing in the Transform Domain	5-4
5.4 Two-Dimensional Discrete Fourier Transform (2D-DFT)	5-5
5.5 Two-Dimensional Discrete Walsh-Hadamard Transformation (2D-DWT)	5-17
5.6 Two-Dimensional Discrete Karhunen-Loève Transform (2D-KLT)	5-24

List of Illustrations

<u>Figure No.</u>	<u>Title</u>	<u>Page No.</u>
3-1	Flow Chart for H_8	3-13
3-2	Walsh Functions	3-15
3-3	Flow Chart of the Cascaded Shuffle Matrix for $N = 8$ Showing that $P_8^3 = I_8$	3-19
3-4a	Flow Chart for $A_8 P_8$	3-21
3-4b	Flow Chart for $P_8 B_8$	3-21
3-5	Haar Functions	3-22
4-1	Spectral Distribution of the Fourier Coefficient $i = 0$	4-9
4-2	Discrete Orthogonal Functions	4-10
4-3	Normalized Autocorrelation Functions	4-19
4-4	Spectral Distribution of Speech Signal and Quantization Error in KLT Processing of 15 Seconds of Speech (case 3 in Table 4-6) ..	4-25
4-5	Autocorrelation Functions, Based on 100 PB Words, for Two Talkers	4-29
4-6	Effect of Removal of $i = 0$ Coefficient	4-35
4-7	rms Error-Rate Distortion Functions for Fourier, Hadamard, Haar, and Karhunen- Loève Transforms	4-40
5-1	Complex Conjugate Symmetry Properties of Two-Dimensional Discrete Fourier Transform ..	5-9
5-2	Two-Dimensional Isovariance Contours in the Fourier Domain	5-11
5-3	Two-Dimensional Walsh-Hadamard Basis for $N = 8$ (black areas represent $+1/N$ and white areas represent $-1/N$)	5-19

PRECEDING PAGE BLANK NOT FILMED

List of Illustrations (Continued)

<u>Figure No.</u>	<u>Title</u>	<u>Page No.</u>
5-4	Normalized Autocorrelation Function of the First-Order Markov Process and Vertical and Horizontal Autocorrelation Functions of the Moonscape	5-29
5-5	Eigenvectors of the Moonscape Covariance Matrix Based on the Horizontal Correlations and Eigenvectors of the First-Order Markov Process as Compared with Harmonic Sinusoidal Functions	5-33
6-1	Signal-to-Quantizing Noise vs Amplitude Loading Factor, α , for Various Bit Assignments	6-5
6-2	PCM Processed Images	6-7
6-3	2D-DFT Processed Images	6-9
6-4	2D-DWT (Hadamard) Processed Images	6-12
6-5	2D-DKLT Processed Images	6-14

List of Tables

<u>Table No.</u>	<u>Title</u>	<u>Page No.</u>
4-1	Test Sentences Used for Experimental Evaluations	4-8
4-2	Fourier Transform (N = 16)--The Standard Deviations and the Optimum Bit Assignment for Fourier Coefficients	4-12
4-3	Hadamard Transform (N = 16)--The Standard Deviations and the Optimum Bit Assignment for Hadamard Coefficients	4-15
4-4	Haar Transform (N = 16)--The Standard Deviations and the Optimum Bit Assignment for Haar Coefficients	4-16
4-5	Karhunen-Loève Transform (N = 16)--Eigenvalues and Computed Variances	4-21
4-6	Karhunen-Loève Transform (N = 16)--Optimum Bit Assignments for KLT Coefficients	4-23
4-7	PB Word Lists	4-26
4-8	Coefficient Bit Assignments for PB Word List Processing	4-30
4-9	Signal-to-Quantizing Noise Ratios for PB Words	4-31
4-10	Assignments of the Number of Bits per Orthogonal Coefficient for Various Speech Processing Schemes	4-33
4-11	Signal-to-Quantizing Noise Ratios for Fourier, Hadamard, Haar, and Karhunen-Loève Transforms	4-41
4-12	PB Word Scores (150 words, two talkers)	4-43
5-1	Optimum Bit Distribution in Two-Dimensional Complex Fourier Domain (average seven bits per sample)	5-14

List of Tables (Continued)

<u>Table No.</u>	<u>Title</u>	<u>Page No.</u>
5-2	Optimum Bit Distribution in Two-Dimensional Complex Fourier Domain (average four bits per sample)	5-15
5-3	Optimum Bit Distribution in Two-Dimensional Complex Fourier Domain (average two bits per sample)	5-16
5-4	Optimum Bit Distribution in Two-Dimensional Walsh-Hadamard Domain (average seven bits per sample)	5-21
5-5	Optimum Bit Distribution in Two-Dimensional Walsh-Hadamard Domain (average four bits per sample)	5-22
5-6	Optimum Bit Distribution in Two-Dimensional Walsh-Hadamard Domain (average two bits per sample)	5-23
5-7	Eigenvalues of the First-Order Markov Process ($\rho = 0.9$)	5-30
5-8	Eigenvalues of the First-Order Markov Process ($\rho = 0.5$)	5-31
5-9	Eigenvalues of Spatial Autocovariance Matrices Representing the Correlations in the x and y Dimensions for Moonscape Data	5-34
5-10	Optimum Bit Distribution in Two-Dimensional Karhunen-Loève Transform Domain (average seven bits per sample)	5-35
5-11	Optimum Bit Distribution in Two-Dimensional Karhunen-Loève Transform Domain (average four bits per sample)	5-36

List of Tables (Continued)

<u>Table No.</u>	<u>Title</u>	<u>Page No.</u>
5-12	Optimum Bit Distribution in Two-Dimensional Karhunen-Loève Transform Domain (average two bits per sample)	5-37
6-1	Optimum Amplitude Loading Factor of the Quantizer in the Walsh-Hadamard Transform Domain	6-6
6-2	Signal-to-Quantizing Noise Ratio Improvement (Λ_Q) of 2D-DFT Over PCM as Applied to Moonscape Slide	6-17
6-3	Signal-to-Quantizing Noise Ratio Improvement of Two-Dimensional Orthogonal Transforms Over PCM as Applied to Moonscape Slide	6-17

1. INTRODUCTION

Many communications systems convert analog input signals into digital signals for processing and transmission. An encoding scheme widely used for the transmission of digital signals is PCM, which has received considerable interest since its invention in 1939. Such encoding generally involves sampling the time signal at a uniform rate and encoding the samples in a binary code which provides an adequate number of quantizing levels, selected in accordance with an appropriate law, to maintain a required signal-to-noise ratio. Such techniques have been applied to voice, image, and other signal waveforms, such as telemetry and biomedical data. In general the method used has been one in which the Nyquist sampling rate is the criterion used for establishing the density of samples in the time domain, and the number of quantization levels is determined by rate-distortion criteria, such as mean-squared error. Other distortion criteria, such as articulation index (AI) for voice, and just noticeable differences (JND) for both voice and image, have also influenced the ultimate transmission rate requirements. The latter criterion, of course, implies the use of subjective criteria for the rate-distortion functions.

Conventional PCM techniques require high bit rates in order to transmit the signals, such as television and speech, with a given level of distortion. Reduction of the bit rate required for the transmission of a speech or television channel is possible through proper coding of the coefficients obtained from a suitable orthogonal transform of the signal. The amount of success achieved thereby depends on the suitability of the transformation for representing the signal. The relative complexity of the implementation with respect to other existing techniques is an important factor that should be considered. The overall

performance can then be judged by using the conventional PCM as a basis for comparison.

Numerous orthogonal transformations for signal processing are conceivable; however, certain ones appear more appropriate than others because the intrinsic form of the members of the orthogonal function set appears to closely resemble the function to be modeled and because of the simplicity of generating the member functions and determining the function coefficients. The orthogonal transforms considered in this study are discrete versions of Fourier, Walsh (Hadamard), Haar, and Karhunen-Loève transformations.

Orthogonal transformation techniques have been suggested by a number of investigators; these techniques are now in various stages of development. A bibliography of works concerning applications to image processing is included as Section 7.6. Kramer and Mathews suggested the use of a linear transformation (based on the eigenvectors of the covariance matrix of a set of signals, i.e., Karhunen-Loève transformation) for transmitting a set of correlated signals, with specific application to information rate reduction of a channel vocoder.¹ Kulya applied such a technique to vocoder parametric control signals in his orthogonal vocoder.² Crowther and Rader have applied the Hadamard transformation to the information rate reduction of vocoder signals.³ Huang and Schultheiss described the development of the discrete version of the linear transformation technique which was the basis for the Karhunen-Loève transform method used in this investigation.⁴

Other techniques such as DPCM are available for accomplishing bit-rate reduction for the transmission of speech and television signals. Linear orthogonal transformation and predictive coding methods are compared in a paper by Nitadori.⁵

In this study, the application of various orthogonal transformations has been investigated, with particular emphasis on speech and visual signal processing. In the sections that follow, the fundamentals of the one- and two-dimensional orthogonal transforms (Sections 3 and 5) and their application to speech and visual signals (Sections 4 and 6) are treated in detail. Summary and conclusions for the speech and visual signal applications are given in Sections 4.10 and 6.7, respectively.

2. RATE-DISTORTION CRITERIA

2.1 GENERAL RATE-DISTORTION CRITERIA

The exact transmission of data from analog sources requires a channel of infinite capacity. Transmission with infinite fidelity is impossible since no physical channel has infinite capacity. If, however, some error or distortion is allowed in the representation of source output by a discrete random variable, then a finite information rate is obtained whose value depends on the amount of distortion allowed.

The relationship between allowable distortion and the maximum rate (in bits per source symbol) which can theoretically be obtained over a channel with capacity C (in bits per second) was first formulated by Shannon as the rate-distortion function, $R(D)$. The rate-distortion function can be interpreted as the minimum average number of bits which may be used for encoding data to be transmitted from a source (which has given probability distributions) with successive, statistically independent symbols without causing the average distortion to exceed D . Multiplying by the number of samples per second, $R(D)$, yields the minimum bit rate required to transmit the source signal subject to the distortion, D . Then, for any source-channel pair, the quantity $R(D) \leq C$ can be used to determine the minimum error achievable for a certain rate without specifying a particular communication system.

Numerous papers have applied the theory of Shannon's rate-distortion function in comparing the performance of various digital transmission systems. A rather complete bibliography is included as Section 7.1. The usefulness of this theory is limited because it is difficult to find accurate mathematical models

for real random processes (such as speech and video signals) and to evaluate the rate-distortion function for any but simple sources. Selection of the distortion measure D is critical in obtaining a meaningful rate-distortion function.

2.2 RATE DISTORTION FOR A DISCRETE MEMORYLESS SOURCE

To gain a further understanding of the rate-distortion function, the rate-distortion function for the discrete memoryless channel will be discussed in the following paragraphs. It should be recognized that this function is not strictly applicable to either the speech or visual channel and is presented only as an example.

The rate-distortion function for a discrete memoryless source is defined as follows by Gallager.⁶ Consider a discrete memoryless source with the alphabet $(0, 1, \dots, K - 1)$ and letter probabilities $Q(0), \dots, Q(K - 1)$, $K < \infty$. The source output is a sequence μ_1, μ_2, \dots , of independent values selected from the source alphabet. The source sequence is represented at the destination by the sequence v_1, v_2, \dots , selected from an alphabet $(0, 1, \dots, J - 1)$, $J < \infty$. A distortion measure $D(j;k)$ is defined for $0 \leq k \leq K - 1$. If source letter k is reconstructed at the receiver as letter j , $0 \leq j \leq J - 1$, which assigns a numerical value to the distortion. The total distortion between the source sequence μ_1, \dots, μ_N and destination sequence v_1, \dots, v_N is simply

$$\sum_{n=1}^N D(\mu_n; v_n)$$

As an example of a distortion measure, let $J = K$ and $D(k;J) = 0$ for $j = k$, and let $J \neq K$ and $D(k;j) = 1$ for $j \neq k$. Such a measure is useful for cases in which all errors are equally serious. Another example is the familiar squared error distortion measure defined by letting $D(k;j) = (j - k)^2$ for all values of j and k .

In order to define the rate-distortion function, consider a set of transition probabilities $P(j/k)$ which are determined by the channel characteristics and all of the processing performed between the source and the destination. $P(j/k)$ is the probability of receiving the letter j at the destination if the source letter was k . These probabilities along with the source probabilities determine the average mutual information:

$$I(Q;P) = \sum_{k=0}^{K-1} \sum_{j=0}^{J-1} Q(k) P(j/k) \ln \frac{P(j/k)}{\sum_i Q(i) P(j/i)}$$

The average distortion is given by

$$\bar{D} = \sum_k \sum_j Q(k) P(j/k) d(k;j)$$

If the maximum allowable value of \bar{D} is equal to D^* (the fidelity criterion), then the rate-distortion function of the source relative to the given distortion measure is defined as

$$R(D^*) = \min_{P: \bar{D} \leq D^*} I(Q;P)$$

where the minimization is over all assignments of transition probabilities subject to the constraint that $\bar{D} \leq D^*$.

Rate-distortion theory is clearly applicable to the familiar problems of analog-to-digital conversion and data rate compression. In practice, the application is difficult because mathematical characterization of important sources such as speech and image signals is not an easy task. Moreover, the formulation of meaningful mathematical distortion measures for such sources has not been accomplished. Despite these complexities, rate-distortion theory does provide some valuable insights into the evaluation of system performance.

2.3 APPLICATION TO SPEECH AND VISUAL SIGNAL PROCESSING

Mean-square error (MSE) is used by numerous investigators because it is relatively simple to compute and sufficiently meaningful. If MSE is taken for D , then, roughly speaking, the function $R_x(D)$ is the minimum rate at which information can be received about signal x to reproduce this information with a mean-square error per symbol equal to D .

The use of the mean-square error is common in the case of speech; it has also been used to evaluate the relative performance of various coding systems, such as PCM and Delta modulation. However, some investigators feel that mean-square error is not entirely satisfactory for speech and have resorted to theoretical computation of the measure known as articulation index (AI) and to subjective assessment in terms of PB words, rhyme tests, or message texts. Articulation index is more difficult to compute than mean-square error, but it has the advantage of being related to sentence and word intelligibility.* For evaluation of

*The method of AI computation is described in fairly good detail in Reference 7, pp.3-5 to 3-12. We have used this method for evaluating the theoretical intelligibility performance of FM, PCM, and DM systems. Results obtained have closely corresponded to those of subjective evaluations.

performance of speech processing methods, it is recommended that computed MSE be used as the principal computed rate-distortion criteria. In order to provide a measure of the effectiveness of types of processing, the MSE of these types of processing will be compared with the MSE computed for conventional companded PCM processing at the same bit rate.

The selection of a suitable distortion measure to evaluate image processing systems is more difficult. It has been shown that simple application of mean-square error as a performance measure for image quality has some limitations. Heuristically, this is easy to see if one considers that errors which affect edges or contours of pictorial data are much more disturbing than errors which affect portions of an image with little or no activity.

The introduction of peak error as a performance criterion for image data may be a better choice, but is again not totally satisfactory. Alternatively, some investigators have suggested that frequency-weighted rms error should be used for PCM processing. However, since a more suitable quantitative distortion measure for image signals has not yet been established, it was decided to employ the MSE criterion for determining the relative performance of each technique in terms of signal-to-noise ratio. It is not expected that the use of other criteria would alter the ranking achieved by using the MSE criterion. As in the case of speech, the effectiveness of the orthogonal processing was gauged by comparing the MSE obtained with the MSE of conventional linear PCM processing. In addition, critical observer tests were performed on the processed image in order to determine the relative image quality produced by the Fourier, Hadamard, and Karhunen-Loève transformations.

3. ORTHOGONAL TRANSFORMATION TECHNIQUES FOR ONE DIMENSION

3.1 GENERAL

As stated previously, numerous orthogonal transformations for signal processing are conceivable; however, certain ones appear more appropriate than others because the intrinsic form of the members of the orthogonal function appears to closely resemble the function to be modeled and because of the simplicity of generating the member functions and determining the function coefficients.

In this study, the Fourier, Hadamard, Haar, and Karhunen-Loève orthogonal transformations are investigated. Except for the Karhunen-Loève transformation, these orthogonal transformations are predetermined in terms of members of the orthogonal function set. The Karhunen-Loève method yields an orthogonal function set that optimally expresses the information content of the signal. Thus, its orthogonal function set is not predetermined, but is instead determined from the autocovariance matrix of the signal to be represented. Consequently, of the methods tested, the Karhunen-Loève method should and does yield the lowest MSE for a given information rate.

Each of the transformations is discussed in detail in the following paragraphs; their application to speech signal processing is discussed in Sections 4 and 5.

3.2 THE DISCRETE FOURIER TRANSFORM (DFT)

The operation of the discrete Fourier transformation may be described as follows:

$$\begin{aligned}\underline{f} &= F_N \underline{s} \\ \underline{s} &= (s_0, s_1, \dots, s_{N-1})^T \\ \underline{f} &= (f_0, f_1, \dots, f_{N-1})^T\end{aligned}\tag{3-1}$$

where s_i ($i = 0, 1, \dots, N - 1$) are the speech samples, f_i represents the Fourier-transformed samples, and F_N is the N th-order discrete Fourier transform matrix.

The discrete Fourier transform $\{f\}$ and its inverse for a real-valued sequence $\{s\}$ of length N are defined* by

$$f_j = \sum_{k=0}^{N-1} s_k e^{-i(2\pi jk)/N}, \quad (j = 0, 1, \dots, N - 1) \tag{3-2}$$

$$s_k = \frac{1}{N} \sum_{j=0}^{N-1} f_j e^{+i(2\pi jk)/N}, \quad (k = 0, 1, \dots, N - 1) \tag{3-3}$$

When W is substituted for $i(2\pi/N)$, equations (3-2) and (3-3) take the familiar form of the discrete Fourier transforms:

$$f_j = \sum_{k=0}^{N-1} s_k W^{-jk} \tag{3-4}$$

*The definitions of the "discrete Fourier transform" are not uniform in the literature. Some authors use f_j/N as the discrete Fourier coefficients, some use $1/\sqrt{N}$ in equations (3-2) and (3-3), and others use a positive exponent in equation (3-2) and a negative exponent in equation (3-3).

$$s_k = \frac{1}{N} \sum_{j=0}^{N-1} f_j w^{+jk} \quad (3-5)$$

It should be noted that, although $\{s\}$ is a real-valued sequence of length N , $\{f\}$ is a complex-valued sequence of the same length. The complex sequence $\{f\}$ can, however, be written in terms of a real-valued sequence $\{f'\}$ of length N by noting that f_0 and $f_{N/2}$ are purely real and that

$$f_{N-j} = f_j^*, \quad (j = 1, 2, \dots, \frac{N}{2} - 1)$$

where f_j^* denotes the complex conjugate of f_j . This statement can be proven simply by using equation (3-4). Since $w^0 = 1$,

$$f_0 = \sum_{k=0}^{N-1} s_k \quad (3-6)$$

and since $w^{-kN/2} = (-1)^k$,

$$f_{N/2} = \sum_{k=0}^{N-1} (-1)^k s_k \quad (3-7)$$

$$f_{N-j} = \sum_{k=0}^{N-1} s_k w^{+jk} \triangleq f_j^* \quad (3-8)$$

and $w^{Nk} = 1$ for all k .

Rewriting equation (3-5) in terms of equations (3-6) and (3-7) results in

$$s_k = \frac{1}{N} \left\{ f_0 + (-1)^k f_{N/2} + \sum_{j=1}^{(N/2)-1} f_j W^{+jk} + \sum_{j=(N/2)+1}^{N-1} f_j W^{+jk} \right\} \quad (3-9)$$

Next, it can be observed that

$$\sum_{j=(N/2)+1}^{N-1} f_j W^{+jk} = \sum_{\ell=1}^{(N/2)-1} f_{N-\ell} W^{+(N-\ell)k}$$

which can be rewritten as

$$\sum_{\ell=1}^{(N/2)-1} f_{\ell}^* W^{+Nk} \cdot W^{-\ell k} = \sum_{\ell=1}^{(N/2)-1} f_{\ell}^* W^{-\ell k}$$

by using equation (3-8). Thus, equation (3-9) can be rewritten as

$$s_k = \frac{1}{N} \left\{ f_0 + (-1)^k f_{N/2} + \sum_{j=1}^{(N/2)-1} \left[f_j W^{+jk} + f_j^* W^{-jk} \right] \right\} \quad (3-10)$$

and

$$s_k = \frac{1}{N} \left\{ f_0 + (-1)^k f_{N/2} + 2 \sum_{j=1}^{(N/2)-1} \operatorname{Re} \left[f_j W^{+jk} \right] \right\} \quad (3-11)$$

where $\operatorname{Re}[\cdot]$ denotes the real part of the complex quantity $f_j W^{+jk}$.

Finally, the real sequence $\{f'\}$ of length N is defined as follows:

$$\left. \begin{aligned} f'_0 &= f_0 \\ f'_{2j-1} &= \operatorname{Re}\{f_j\} \\ f'_{2j} &= -\operatorname{Im}\{f_j\} \\ f'_{N-1} &= f_{N/2} \end{aligned} \right\} \quad j = (1, 2, \dots, \frac{N}{2} - 1) \quad (3-12)$$

The inverse Fourier transform can then be written in terms of the sequence $\{f'\}$ as

$$s_k = \frac{1}{N} \left\{ f'_0 + (-1)^k f'_{N-1} + \sum_{j=1}^{(N/2)-1} \left[f'_{2j-1} \cos(\delta j k) + f'_{2j} \sin(\delta j k) \right] \right\} \quad (3-13)$$

where $\delta = 2\pi/N$.

If the sequence $\{f'\}$ is transmitted instead of the signal sequence $\{s\}$, the signal can be reconstructed at the receiving side by applying the inverse Fourier transformation of equation (3-13) to $\{f'\}$. The reconstructed signal $\{s\}$ differs from the original signal $\{s\}$ because of the quantization errors and the error caused by the channel. Because of the averaging property of the transformation, it may be suspected that transformation of the Fourier-transformed speech instead of the speech signal itself provides enhanced toleration to channel errors. It will be shown that the signal-to-quantization noise ratio is also improved.

The fast Fourier transform algorithm is an efficient method for computing the discrete Fourier transform,

$$f_j = \sum_{k=0}^{N-1} s_k \exp(-i2\pi jk/N) \quad (3-14)$$

for $j = 0, 1, \dots, N - 1$, where $\{f\}$ and $\{s\}$ are generally complex-valued. The basic idea is to factor N , i.e.,

$$N = \prod_{i=1}^m N_i$$

and to then decompose the transform into m steps with N/N_i transformations of size N_i within each step. This idea was first proposed by Cooley and Tukey.⁸

Numerous subroutines based upon modifications of this algorithm have been published in recent years. Most of these have dealt with the special case $N = 2^n$. The computation time required by a well-written subroutine based upon this algorithm is approximately proportional to $N \sum N_i$. If equation (3-14) were programmed directly, the required computation time would be proportional to N^2 . Since, for large N , $\sum N_i \ll N$, the computation time is reduced considerably. For example, if $N = 1024$, the fast Fourier transform can be computed in about two percent of the time required by a direct computation.

The subroutine used here is that of R. C. Singleton.⁹ It is based upon a mixed-radix algorithm which makes it possible to compute the discrete Fourier transform for any arbitrary length sequence; i.e., N is not restricted to 2^n . Because the FFT is discussed thoroughly in Reference 9, no further discussion is included here.

3.3 THE DISCRETE WALSH-HADAMARD TRANSFORM (DWT)

3.3.1 Hadamard Matrices

The Hadamard matrix is a square array of plus and minus 1's whose rows (and columns) are orthogonal to one another.^{10, 11, 12} If H_N is a symmetrical Hadamard matrix of order N , then

$$H_N H_N = N I_N \quad (3-15)$$

where I_N is the identity matrix of dimension N . In a symmetrical Hadamard matrix, it is possible to

- (a) interchange rows,
- (b) interchange columns,
- (c) change the sign of every element in a row, or
- (d) change the sign of every element in a column

without affecting the orthogonality properties of the matrix. These operations make it possible to achieve a symmetrical Hadamard matrix whose first row and first column contain only +1's. The matrix thus obtained is called "normal form" for the Hadamard matrix and is equivalent to all the other matrices that can be obtained from it by using these basic operations. However, it should be noted that the normal form is not unique within an equivalence class. Further, there is more than one equivalence class of Hadamard matrices of a given dimension N , $N \geq 16$.

The lowest order Hadamard matrix is of the order two;
i.e.,

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

The existence of Hadamard matrices of all possible dimensions is an unsolved problem in mathematics. However, it is known that, if a Hadamard matrix of order N exists ($N > 2$), then $N \equiv (\text{Mod } 4)$; i.e., $N = 4, 8, 12, 16, \dots$. The construction of high-order Hadamard matrices is easier if the order is restricted further to powers of two ($N = 2^n$). From now on, only symmetrical Hadamard matrices in normal form and of dimension $N = 2, 4, 8, 16, 32, 64, \dots$, will be considered in this report. These high-order Hadamard matrices can be generated recursively by using Kronecker products.^{*,10,12,13} The recursion starts with H_2 . In the

*For two matrices $A(a_{ij})$ and B , the Kronecker product $A \times B$ is defined as: $A \times B = (a_{ij}B)$. For example, if

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \quad B = \begin{bmatrix} b_{11} & b_{12} \\ b_{12} & b_{22} \\ b_{13} & b_{23} \end{bmatrix}$$

then

$$A \times B = \begin{bmatrix} a_{11}b_{12} & a_{12}b_{11} & a_{11}b_{12} & a_{12}b_{12} \\ a_{21}b_{11} & a_{22}b_{11} & a_{21}b_{12} & a_{22}b_{12} \\ a_{11}b_{12} & a_{12}b_{12} & a_{11}b_{22} & a_{12}b_{22} \\ a_{21}b_{12} & a_{22}b_{12} & a_{21}b_{22} & a_{22}b_{22} \\ a_{11}b_{13} & a_{12}b_{13} & a_{11}b_{23} & a_{12}b_{23} \\ a_{21}b_{13} & a_{22}b_{13} & a_{21}b_{23} & a_{22}b_{23} \end{bmatrix}$$

(continued on next page)

following matrices, the elements of the matrices will be denoted only as \pm instead of ± 1 .

$$H_2 = \begin{pmatrix} + & + \\ + & - \end{pmatrix} \quad \begin{array}{l} \text{No. of Sign} \\ \text{Changes} \end{array} \quad \begin{array}{l} 0 \\ 1 \end{array} \quad (3-16)$$

$$H_4 = H_2 \times H_2 = \begin{pmatrix} H_2 & H_2 \\ H_2 & H_2 \end{pmatrix} = \begin{pmatrix} + & + & + & + \\ + & - & + & - \\ + & + & - & - \\ + & - & - & + \end{pmatrix} \quad \begin{array}{l} \text{No. of Sign} \\ \text{Changes} \end{array} \quad \begin{array}{l} 0 \\ 3 \\ 1 \\ 2 \end{array} \quad \begin{array}{l} 00 \\ 11 \\ 01 \\ 10 \end{array} \quad (3-17)$$

$$H_8 = H_2 \times H_4 = \begin{bmatrix} + & + & + & + & + & + & + & + \\ + & - & + & - & + & - & + & - \\ + & + & - & - & + & + & - & - \\ + & - & - & + & + & - & - & + \\ + & + & + & + & - & - & - & - \\ + & - & + & - & - & + & - & + \\ + & + & - & - & - & - & + & + \\ + & - & - & + & - & + & + & - \end{bmatrix} \quad \begin{array}{l} \text{No. of Sign} \\ \text{Changes} \end{array} \quad \begin{array}{l} 0 \\ 7 \\ 3 \\ 4 \\ 1 \\ 6 \\ 2 \\ 5 \end{array} \quad \begin{array}{l} 000 \\ 111 \\ 011 \\ 100 \\ 001 \\ 110 \\ 010 \\ 101 \end{array} \quad (3-18)$$

The Kronecker product can be combined with matrix multiplication through the formula,

$$(A \times B)(C \times D) = (AC) \times (BD)$$

if the dimensions of the matrices are compatible.

In general, the Nth-order Hadamard matrix is generated from $H_{N/2}$ as

$$H_N = H_2 \times H_{N/2} = \begin{pmatrix} H_{N/2} & H_{N/2} \\ H_{N/2} & -H_{N/2} \end{pmatrix} \quad (3-19)$$

which can be written as

$$H_N = \prod_{i=1}^n (I_{2^{n-1}} \times H_2 \times H_{2^{i-1}}) = \prod_{i=1}^n h_i \quad (3-20)$$

from References 12 and 13, where

$$n = \log_2 N \quad (\text{or } N = 2^n)$$

$$h_i = I_{2^{n-1}} \times H_2 \times I_{2^{i-1}}, \quad i = 1, 2, 3, \dots, n$$

$$I_k \triangleq k \times k = \text{identity matrix}$$

The proof of equation (3-20) is done by iteration.
Note that H_N in equation (3-19) can be written as

$$H_N = \begin{pmatrix} H_{N/2} & 0 \\ 0 & H_{N/2} \end{pmatrix} \begin{pmatrix} I_{N/2} & I_{N/2} \\ I_{N/2} & I_{N/2} \end{pmatrix} = (I_2 \times H_{N/2}) (H_2 \times I_{N/2}) \quad (3-21)$$

where

$$\begin{aligned} H_{N/2} &= \begin{pmatrix} H_{N/4} & H_{N/4} \\ H_{N/4} & -H_{N/4} \end{pmatrix} = \begin{pmatrix} H_{N/4} & 0 \\ 0 & H_{N/4} \end{pmatrix} \begin{pmatrix} I_{N/4} & I_{N/4} \\ I_{N/4} & -I_{N/4} \end{pmatrix} \\ &= (I_2 \times H_{N/4}) (H_2 \times I_{N/4}) \end{aligned}$$

Equation (3-21) can be written as

$$\begin{aligned} H_N &= [I_2 \times (I_2 \times H_{N/4}) (H_2 \times I_{N/4})] (H_2 \times I_{N/2}) \\ &= (I_4 \times H_{N/4}) (I_2 \times H_2 \times I_{N/4}) (H_2 \times I_{N/2}) \end{aligned} \quad (3-22)$$

where

$$H_{N/4} = (I_2 \times H_{N/8}) (H_2 \times I_{N/8})$$

and equation (3-22) can be written as

$$\begin{aligned} H_N &= (I_8 \times H_{N/8}) (I_4 \times H_2 \times I_{N/8}) (I_2 \times H_2 \times I_{N/4}) \\ &\quad \times (H_2 \times I_{N/2}) \end{aligned} \quad (3-23)$$

If we continue in the same manner with

$$\begin{aligned} I_{N/4} \times H_4 &= I_{N/4} \times (I_2 \times H_2) (H_2 \times I_2) \\ &= (I_{N/2} \times H_2) (I_{N/4} \times H_2 \times I_2) \end{aligned}$$

equation (3-23) can be further expanded as

$$\begin{aligned} H_N &= \underbrace{(I_{N/2} \times H_2)}_{h_1} \underbrace{(I_{N/4} \times H_2 \times I_2)}_{h_2} \underbrace{(I_{N/8} \times H_2 \times I_4)}_{h_3} \dots \\ &\quad \times (I_4 \times H_2 \times I_{N/8}) \underbrace{(I_2 \times H_2 \times I_{N/4})}_{h_{n-1}} \underbrace{(H_2 \times I_{N/2})}_{h_n} \end{aligned} \quad (3-24)$$

where the 1st, 2nd, 3rd, . . . , nth terms are the $h_1, h_2, h_3, \dots, h_n$ of equation (3-20), respectively. Equation (3-20) is thus a compact form of equation (3-24).

Further examination of equation (3-20) shows that it is the product of $n = \log_2 N$ terms, which are symmetric square matrices of dimension N . Any row (or column) of these matrices has only two non-zero entries of either +1 or -1.* There are N add operations in each matrix, giving $N \log_2 N$ algebraic operations.

In Section 3.3.2 a different form of equation (3-20) will be derived. It will be shown that, although the algorithm for this new form does not result in a reduction of the number of basic operations, it is especially suitable for programming and implementation purposes. Before the new algorithm is explained, however, sequency should be considered. Sequency (Reference 14, p. 50) is closely related to the number of sign changes along each row of the Hadamard matrix. If the number of sign changes along a row is i , the sequency of the Walsh function¹⁵ corresponding to that row is given by

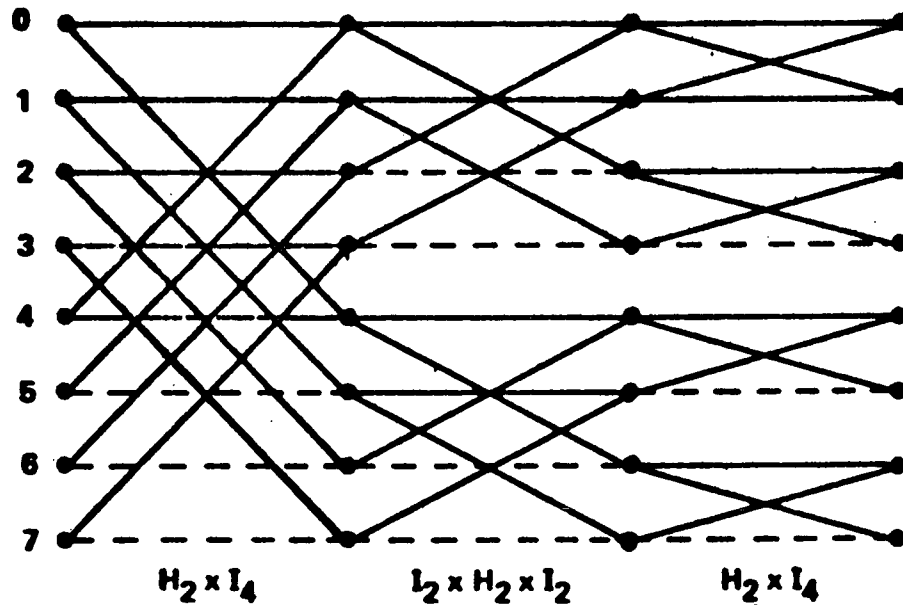
$$j = \begin{cases} i/2 & \text{if } i \text{ is even} \\ \frac{i+1}{2} & \text{if } i \text{ is odd} \end{cases}$$

*For example, for $N = 8$, $n = 3$, equation (3-24) yields

$$H_8 = (I_4 \times H_2)(I_2 \times H_2 \times I_2)(H_2 \times I_4)$$

$$= \begin{bmatrix} + & + & . & . & . & . & . & . \\ + & - & . & . & . & . & . & . \\ . & . & + & + & . & . & . & . \\ . & . & + & - & . & . & . & . \\ . & . & . & . & + & + & . & . \\ . & . & . & . & + & - & . & . \\ . & . & . & . & . & . & + & + \\ . & . & . & . & . & . & + & - \end{bmatrix} \begin{bmatrix} + & . & + & . & . & . & . & . \\ . & + & . & + & . & . & . & . \\ + & . & - & . & . & . & . & . \\ . & + & . & - & . & . & . & . \\ . & . & . & . & + & + & . & . \\ . & . & . & . & . & + & + & . \\ . & . & . & . & + & . & - & . \\ . & . & . & . & . & + & . & - \end{bmatrix} \begin{bmatrix} + & . & . & . & + & . & . & . \\ . & + & . & . & . & + & . & . \\ . & . & + & . & . & . & + & . \\ . & . & . & + & . & . & . & + \\ + & . & . & . & - & . & . & . \\ . & + & . & . & . & - & . & . \\ . & . & + & . & . & . & - & . \\ . & . & . & + & . & . & . & - \end{bmatrix}$$

which, when multiplied, gives H_8 of equation (3-18) (Figure 3-1 is the flow chart for H_8 .)

Figure 3-1. Flow Chart for H_8

The rows of the Hadamard matrices developed by using equation (3-20) are in "natural" order. This is shown for H_2 , H_4 , and H_8 in equations (3-16) through (3-18). The number of sign changes in any row is given in decimal and binary form on the right of the matrices. The rows can be rearranged to achieve sequency ordering. Such a Hadamard matrix would be denoted as H_N^W , where the superscript W denotes Walsh. As an example, the Walsh functions corresponding to the rows of the matrix H_8^W will be given next:

		No. of Sign Changes (i)	Se- quency (j)	Walsh Functions
$H_8^W =$	$\begin{bmatrix} + & + & + & + & + & + & + & + \\ + & + & + & + & - & - & - & - \\ + & + & - & - & - & - & + & + \\ + & + & - & - & + & + & - & - \\ + & - & - & + & + & - & - & + \\ + & - & - & + & - & + & + & - \\ + & - & + & - & - & + & - & + \\ + & - & + & - & + & - & + & - \end{bmatrix}$	0	0	wal(0,x) or cal(0,x)
		1	1	wal(1,x) or sal(1,x)
		2	1	wal(2,x) or cal(1,x)
		3	2	wal(3,x) or sal(2,x)
		4	2	wal(4,x) or cal(2,x)
		5	3	wal(5,x) or sal(3,x)
		6	3	wal(6,x) or cal(3,x)
		7	4	wal(7,x) or sal(4,x)

The functions wal(i,x) are orthonormal in the interval $0 \leq x \leq 1$ and can be generated by a difference equation or as the product of linear functions. These functions can be divided as even functions [cal(j,x)], odd functions [sal(j,x)], and as the constant 1 or wal(0,x). The first eight Walsh functions are shown in Figure 3-2. The $N \times N$ Hadamard matrix defines N Walsh functions: $j = 0, 1, 2, \dots, N - 1$. The corresponding sal and cal functions are

$$\begin{aligned} \text{cal}(j,x) &= \text{wal}(2j,x) \quad ; \quad j = 0, 1, 2, \dots, (N/2 - 1) \\ \text{sal}(j,x) &= \text{wal}(2j-1,x), \quad j = 1, 2, \dots, N/2 \end{aligned}$$

Reference 11 gives a computational algorithm for an ordered Hadamard transform; this algorithm¹⁶ has the same number of operations ($N \log_2 N$) as any other "fast" Hadamard transform¹² and yields the Hadamard matrix in natural order.

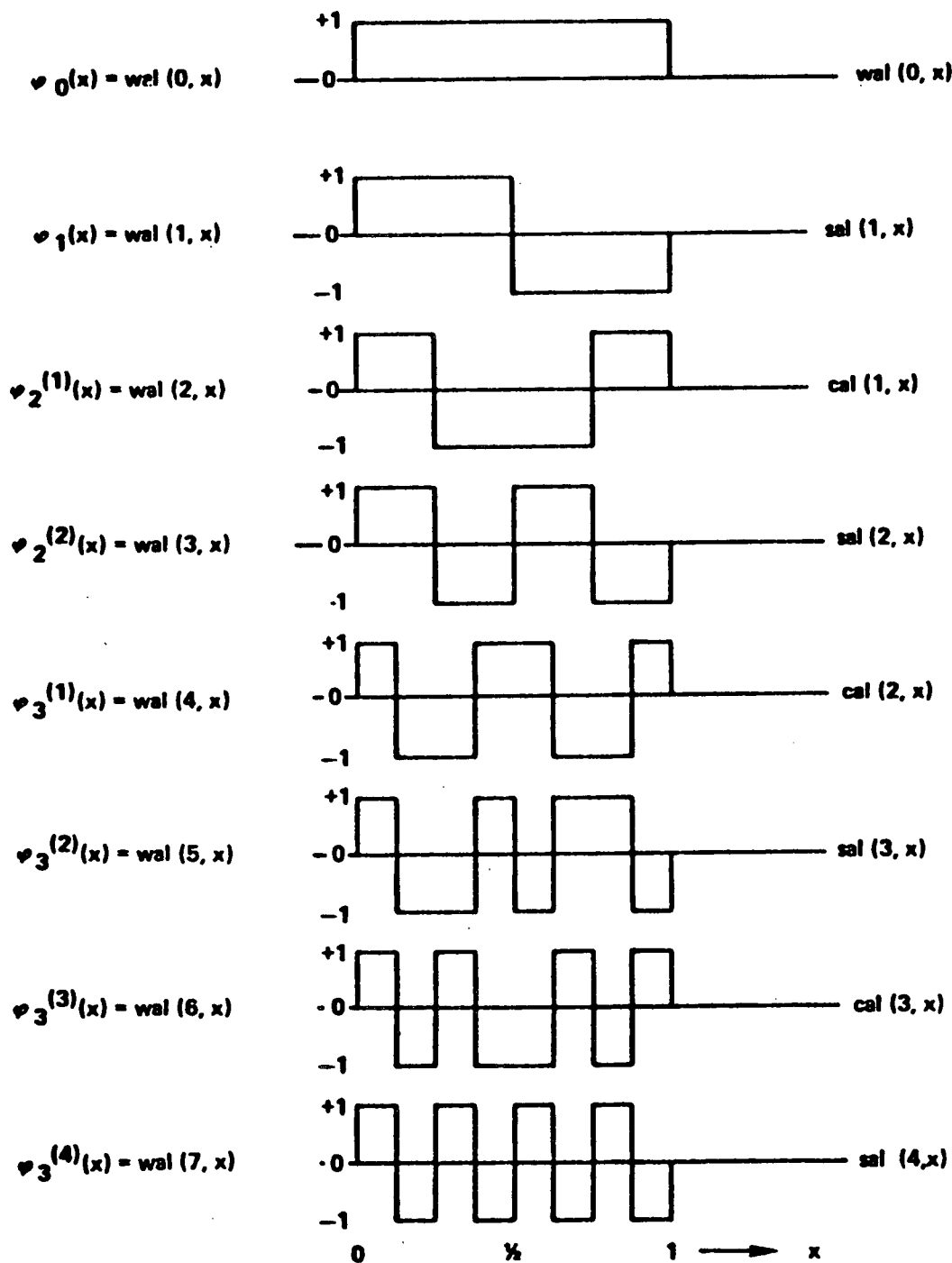


Figure 3-2. Walsh Functions

3.3.2 A Computational Algorithm for Hadamard Transform¹⁶

For the purpose of parallelism, the difficulty with equation (3-20) is that H_2 occurs in different locations in the various terms of the product. Therefore, it must be determined whether there is a permutation operator, P , such that

$$P_N^{-1} = P_N^T \quad (3-25a)$$

$$P_N(I_2 \times Q_{N/2}) P_N^{-1} = Q_{N/2} \times I_2 \quad (3-25b)$$

or

$$P_N^{-1}(Q_{N/2} \times I_2) P_N = I_2 \times Q_{N/2} \quad (3-25c)$$

$$P_N^n = P_N^{-n} = I_N \quad (3-25d)$$

where $Q_{N/2}$ is a square matrix of dimension $N/2$, and $n = \log_2 N$. The behavior of the product operator in an arbitrary vector indicates that a possible choice for P is the "ideal shuffle" matrix.¹³ The particular type of ideal shuffle matrix used in the algorithm is defined by

$$\begin{aligned} P_N &\cdot \text{col}(x_0, x_1, \dots, x_{(N/2)-1}, x_{N/2}, \dots, x_{N-1}) \\ &= \text{col}(x_0, x_{N/2}, x_1, x_{(N/2)+1}, \dots, x_{(N/2)-1}, x_{N-1}) \end{aligned} \quad (3-26)$$

where x_0, x_1, \dots, x_{N-1} are the columns of the N th-order Hadamard matrix in natural order.

Theorem: If P_N is an "ideal shuffle" permutation matrix which satisfies the conditions given by equation (3-25), then equation (3-20) takes the form

$$H_N = (A_N P_N)^n = (P_N B_N)^n \quad (3-27)$$

where

$$A_N = I_{N/2} \times H_2$$

$$B_N = H_2 \times I_{N/2}$$

Proof: Let $Q_{N/2} = I_{N/4} \times H_2$ and note that

$$I_2 \times Q_{N/2} = I_{N/2} \times H_2 = A_N \quad (3-28)$$

Next, it will be shown that the matrices A , $(P_N A P_N^{-1})$, $(P_N A P_N^{-2})$, \dots , $(P_N^{n-1} A P_N^{-(n-1)})$ are equal to the 1st, 2nd, 3rd, \dots , nth terms of equation (3-24), respectively. From equation (3-28),

$$A_N = I_{N/2} \times H_2 = h_1 \quad (3-29a)$$

Using equation (3-29a) and (3-25b) results in

$$\begin{aligned} P_N A_N P_N^{-1} &= P_N (I_2 \times Q_{N/2}) P_N^{-1} = Q_{N/2} \times I_2 = I_{N/4} \times H_2 \times I_2 = h_2 \\ P_N^2 A_N P_N^{-2} &= P_N (P_N A_N P_N^{-1}) P_N^{-1} \end{aligned} \quad (3-29b)$$

From equation (3-29b),

$$P_N^2 A_N P_N^{-2} = P_N (Q_{N/2} \times I_2) P_N^{-1} = P_N (I_{N/4} \times H_2 \times I_2) P_N^{-1}$$

which can be rewritten as

$$P_N^2 A_N P_N^{-2} = P_N (I_2 \times Q_{N/2}^1) P_N^{-1}$$

where

$$Q_{N/2}^1 = I_{N/8} \times H_2 \times I_2$$

is also a square matrix. Reapplying equation (3-25b) results in

$$P_N^2 A_N P_N^{-2} = Q_{N/2}^1 \times I_2 = I_{N/8} \times H_2 \times I_4 = h_3 \quad (3-29c)$$

Finally,

$$P_N^{(n-1)} A_N P_N^{-(n-1)} = H_2 \times I_{N/2} = h_n \quad (3-29d)$$

These results can be used to rewrite equation (3-24) as

$$\begin{aligned} H_N &= h_1 \cdot h_2 \cdot \dots \cdot h_n \\ &= A_N \cdot P_N A_N P_N^{-1} \cdot P_N^2 A_N P_N^{-2} \cdot \dots \cdot P_N^{(n-2)} A_N P_N^{-(n-2)} \\ &\quad \cdot P_N^{(n-1)} A_N P_N^{-(n-1)} \\ &= (A_N P_N)^n \cdot P_N^{-n} \end{aligned}$$

where, according to equation (3-25d), $P_N^{-n} = I_N$; hence, equation (3-27) is obtained. The second form of equation (3-27) can be proven similarly by using B_N (instead of A_N) and the property of P_N given by equation (3-25c).

Examples of "ideal shuffle" permutation matrices are

$$P_4 = \begin{bmatrix} + & \cdot & \cdot & \cdot \\ \cdot & \cdot & + & \cdot \\ \cdot & + & \cdot & \cdot \\ \cdot & \cdot & \cdot & + \end{bmatrix} \quad P_8 = \begin{bmatrix} + & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & + & \cdot & \cdot & \cdot \\ \cdot & + & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & + & \cdot & \cdot \\ \cdot & \cdot & + & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & + & \cdot \\ \cdot & \cdot & \cdot & + & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & + \end{bmatrix} \quad (3-30)$$

where + denotes +1 and, as usual, the dots denote zero entries. The flow chart of P_8^3 (Figure 3-3) shows that $P_8^3 = I_8$.

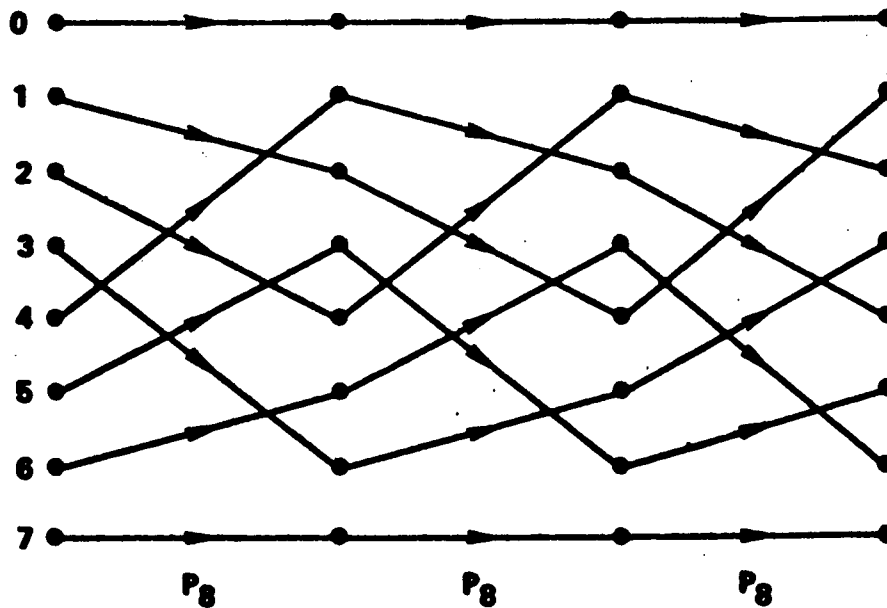


Figure 3-3. Flow Chart of the Cascaded Shuffle Matrix for $N = 8$ Showing that $P_8^3 = I_8$

For $N = 8$,

$$A_8 = I_4 \times H_2 = \begin{bmatrix} + & + & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ + & - & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & + & + & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & + & - & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & + & + & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & + & - & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & + & + \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & + & - \end{bmatrix}$$

$$A_8 = H_2 \times I_4 = \begin{bmatrix} + & \cdot & \cdot & \cdot & + & \cdot & \cdot & \cdot \\ \cdot & + & \cdot & \cdot & \cdot & + & \cdot & \cdot \\ \cdot & \cdot & + & \cdot & \cdot & \cdot & + & \cdot \\ \cdot & \cdot & \cdot & + & \cdot & \cdot & \cdot & + \\ + & \cdot & \cdot & \cdot & - & \cdot & \cdot & \cdot \\ \cdot & + & \cdot & \cdot & \cdot & - & \cdot & \cdot \\ \cdot & \cdot & + & \cdot & \cdot & \cdot & - & \cdot \\ \cdot & \cdot & \cdot & + & \cdot & \cdot & \cdot & - \end{bmatrix}$$

$$A_8 P_8 = P_8 B_8 = \begin{bmatrix} + & \cdot & \cdot & \cdot & + & \cdot & \cdot & \cdot \\ + & \cdot & \cdot & \cdot & - & \cdot & \cdot & \cdot \\ \cdot & + & \cdot & \cdot & \cdot & + & \cdot & \cdot \\ \cdot & + & \cdot & \cdot & \cdot & - & \cdot & \cdot \\ \cdot & \cdot & + & \cdot & \cdot & \cdot & + & \cdot \\ \cdot & \cdot & + & \cdot & \cdot & \cdot & - & \cdot \\ \cdot & \cdot & \cdot & + & \cdot & \cdot & \cdot & + \\ \cdot & \cdot & \cdot & + & \cdot & \cdot & \cdot & - \end{bmatrix}$$

When computed, $(A_8 P_8)^3$ and $(P_8 B_8)^3$ yield H_8 of equation (3-19).
Figures 3-4a and 3-4b are the flow charts for $A_8 P_8$ and $P_8 B_8$.

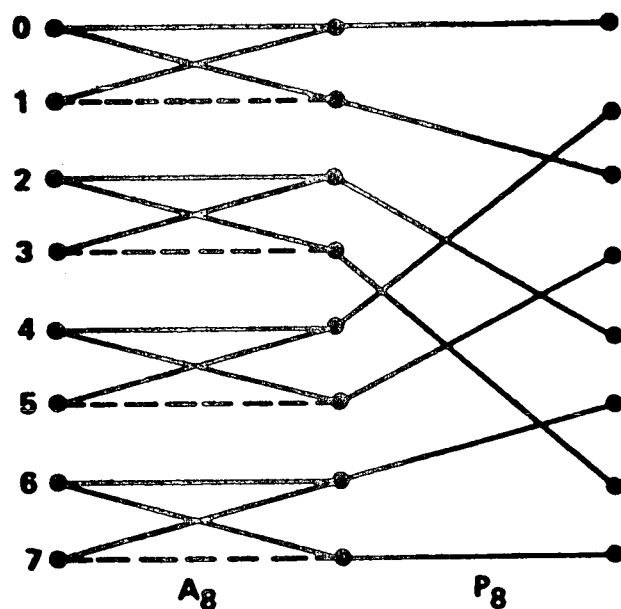


Figure 3-4a. Flow Chart for $A_8 P_8$

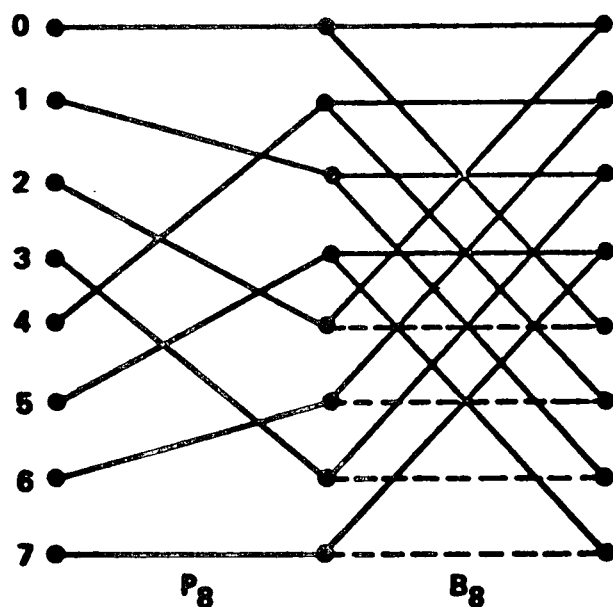


Figure 3-4b. Flow Chart for $P_8 B_8$

3.4 THE DISCRETE HAAR TRANSFORM (DHT)

3.4.1 Haar Functions

Haar functions¹⁷⁻²⁰ are a complete set of orthonormal periodic functions defined on the interval $[0,1]$. A few of the lowest order Haar functions are shown in Figure 3-5.

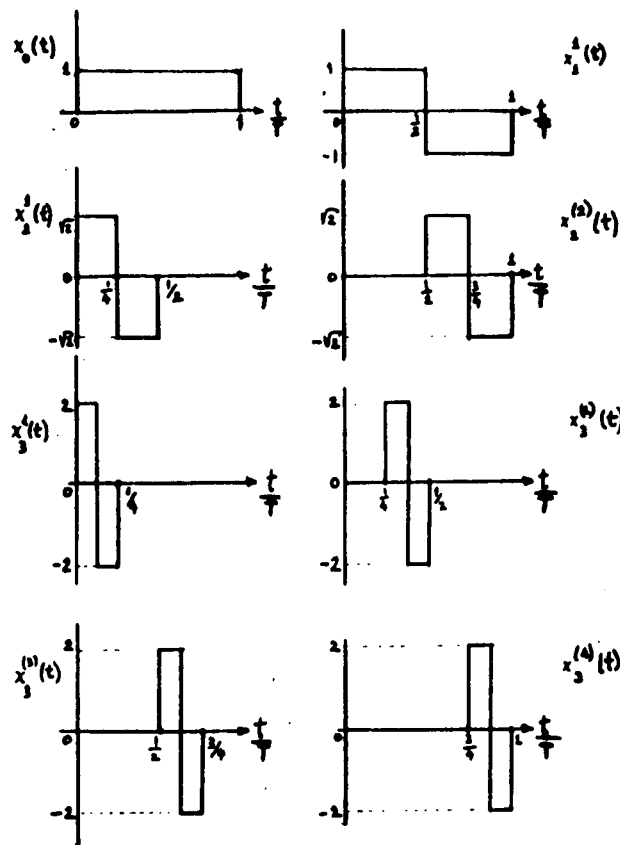


Figure 3-5. Haar Functions

Haar functions can be represented by the orthonormal Haar matrix. As an example, the following orthonormal 8×8 Haar matrix, where the dots represent zeros, is presented.

$$H_8 = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & . & . & . & . \\ . & . & . & . & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & -2 & . & . & . & . & . & . \\ . & . & 2 & -2 & . & . & . & . \\ . & . & . & . & 2 & -2 & . & . \\ . & . & . & . & . & . & 2 & -2 \end{bmatrix} \begin{matrix} i = 1 \\ i = 2 \\ i = 3 \\ \\ i = 4 \end{matrix} \quad (3-31)$$

The Haar coefficients $[a_h(0), a_h(1), \dots, a_h(N-1)]$ corresponding to a signal vector $\underline{s} = [s_0, s_1, \dots, s_{N-1}]$ can be computed from the following matrix relationship:

$$\underline{a}_h = \frac{1}{N} H_N \underline{s} \quad (3-32)$$

where \underline{a}_h is the Haar coefficient vector, H_N is the $N \times N$ Haar transformation matrix, and N is a power of two; i.e., $N = 2^n$, $n = 1, 2, 3, 4, \dots$. It should be noted that the first coefficient, $a_h(0)$, is the average value of N signal samples. The coefficient $a_h(1)$ is the same as the $\text{sal}(1, x)$, $0 \leq x \leq 1$, coefficient in Walsh function analysis.¹⁶ The rest of the coefficients do not bear such a one-to-one resemblance to the Walsh or Fourier coefficients, but they can be obtained as a linear combination of these coefficients.

The $N \times N$ Haar matrix can be written by inspection of the H_8 or directly from the definition¹⁵ of Haar functions:

$$\chi_0(x) = 1, \quad 0 \leq x \leq 1$$

$$\chi_1(x) = \begin{cases} 1, & 0 \leq x < 1/2 \\ -1, & 1/2 < x \leq 1 \end{cases}$$

$$\chi_n^{(k)}(x) = \begin{cases} \sqrt{2^{n-1}}, & \frac{k-1}{2^{n-1}} < x < \frac{2k-1}{2^n}, \quad k = 1, 2, 3, \dots, 2^{n-1} \\ -\sqrt{2^{n-1}}, & \frac{2k-1}{2^n} < x < \frac{k}{2^{n-1}}, \quad n = 1, 2, 3, \dots, \infty \\ 0 & 0 < x < \frac{k-1}{2^{n-1}} \text{ or } \frac{k}{2^{n-1}} < x < 1 \end{cases}$$

The following matrices are 4×4 and 16×16 Haar matrices, respectively. Again, the dots represent zeros and $c = 2\sqrt{2}$.

$$H_4 = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ \sqrt{2} & -\sqrt{2} & . & . \\ . & . & \sqrt{2} & \sqrt{2} \end{bmatrix}$$

$$H_{16} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 & -1 \\ \sqrt{2} & \sqrt{2} & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & -\sqrt{2} & -\sqrt{2} & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & \sqrt{2} & \sqrt{2} & \sqrt{2} & \sqrt{2} & -\sqrt{2} & -\sqrt{2} & -\sqrt{2} & -\sqrt{2} \\ 2 & 2 & -2 & -2 & . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & 2 & 2 & -2 & -2 & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & 2 & 2 & -2 & -2 & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . & 2 & 2 & -2 & -2 \\ c & -c & . & . & . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & c & -c & . & . & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & c & -c & . & . & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & c & -c & . & . & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & c & -c & . & . & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & c & -c & . & . & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . & c & -c & . & . \\ . & . & . & . & . & . & . & . & . & . & . & . & . & c & -c \end{bmatrix}$$

It should be observed that the first row in H_N results in the average value of N samples, and that this value is the same in Walsh-Hadamard or Fourier expansions. The rest of the rows can be divided into $\log_2 N$ groups, as shown for H_8 . This observation is used in developing the Haar transform algorithm.

3.4.2 Haar Transform Algorithm

Andrews and Caspari²¹ have given a factorization of H_8 . A different algorithm for the computation of Haar coefficients in Fortran IV follows. In this program, X and Y are the input and output arrays, respectively, and $\text{LOGN} = \log_2 N$.

```

      SUBROUTINE RHAAR(X,Y,LOGN)
      DIMENSION X(1),Y(1)
      N= 2**LOGN
      SUM= 0.
      DO 10 I=1,N
10      SUM= SUM+X(I)
      Y(1)= SUM/FLOAT(N)
      DO 30 I=1,LOGN
      NJ= 2** (I-1)
      AI= SQRT(FLOAT(NJ))
      DN= AI/FLOAT(N)
      NI= 2** (LOGN-I+1)
      NH1= NI/2
      NH1= NH1-1
      DO 30 J=1,NJ
      J1= J-1
      L= NJ+J
      KD= J1*NI
      SUM= X(1+KD)-X(NI+KD)
      IF (NH1.EQ.0) GO TO 30
      DO 20 K=1,NH1
20      SUM= SUM+X(K+1+KD)-X(K+NH1+KD)
30      Y(L)= SUM*DN
      RETURN
      END

```

The input is the array of N consecutive time samples of the signal $x(t)$, $i = 1, 2, \dots, N$. The output is N Haar coefficients; i.e., $a_h(0), a_h(1), \dots, a_h(N-1)$ and the square magnitudes of the Haar coefficients are $a_h^2(i)$, where $i = 0, 1, \dots, N-1$.

3.4.3 Inverse Haar Transformation

The $N \times N$ Haar matrix, H_N , and its transpose, H_N^T , satisfy the following relationship:

$$H_N H_N^T = N I_N \quad (3-34)$$

where I_N is the $N \times N$ identity matrix. This means that the inverse Haar matrix is equal to the transpose of the Haar matrix except for a factor of $1/N$. That is,

$$H_N^{-1} = \frac{1}{N} H_N^T$$

For example, it can be seen that, for H_4 in Section 3.4.1, the inverse is

$$H_4^{-1} = \frac{1}{4} \begin{bmatrix} 1 & 1 & \sqrt{2} & \cdot \\ 1 & 1 & -\sqrt{2} & \cdot \\ 1 & -1 & \cdot & \sqrt{2} \\ 1 & -1 & \cdot & \sqrt{2} \end{bmatrix}$$

For the Haar coefficients $a_h(i)$ ($i = 0, 1, \dots, N$) as defined by equation (3-32), the discrete time function, $\hat{s}(i)$, can be computed as

$$\underline{\hat{s}} = H_N^T \underline{a}_h \quad (3-35)$$

Computation of the inverse Haar transform from this expression requires N^2 operations. However, grouping the columns of the transposed Haar matrix makes it possible to reduce the number of operations by

$$\frac{\log_2 N}{N}$$

The following program is the Fortran IV program for the inverse Haar transformation. X and Y are the transform and inverse transform arrays, respectively.

```

      SUBROUTINE RIHAAR(X,Y,LOGN)
      DIMENSION X(1),Y(1)
      N= 2**LOGN
      L= 1
      DO 5 I=1,N
5      Y(I)= X(I)
      DO 10 I=1,LOGN
      NJ= 2**((I-1))
      AI= SQRT(FLOAT(NJ))
      NI= 2**((LOGN-I+1))
      NHI= NI/2
      DO 10 J=1,NJ
      J1= J-1
      L= L+1
      KD= J1*NI
      S= AI*X(L)
      DO 10 K=1,NI
      KL= K+KD
      SS= S
10      IF(K.GT.NHI) SS= -S
      Y(KL)= Y(KL)+SS
      RETURN
      END

```

3.5 THE DISCRETE KARHUNEN-LOÈVE TRANSFORMATION--THE OPTIMUM ORTHOGONAL TRANSFORMATION

3.5.1 General

The Karhunen-Loève transformation has long been used as a theoretical tool in the study of random processes.^{1,4,22-24} More recently, a discrete version of the Karhunen-Loève expansion has proved to be a practical tool.²⁵ This finite expansion has found application in pattern recognition and communication theory.^{12,26-30} By taking only a finite number of terms of an infinite series expansion when a random process is expanded in terms of a complete set of orthogonal functions, it minimizes the average error committed.

3.5.2 Karhunen-Loève Transform Development

Assume that a source output waveform $x(t)$ is sampled at regular time intervals. The samples of $x(t)$, designated as $\{x_i\}$, are processed in blocks of N samples. Let X denote a vector of N samples which is transformed to another vector Y by the linear transformation

$$Y = KX \quad (3-36)$$

The components of the vector Y are quantized and transmitted over a noise-free digital channel. At the receiving side the operation is reversed; the signal vector is obtained by the inverse transformation

$$\hat{X} = K^{-1}\hat{Y} \quad (3-37)$$

and then a reasonable facsimile of $x(t)$ is reconstructed at the receiver.

The objective of this procedure is to exploit correlations between input samples $\{x_i\}$ by first generating a set of uncorrelated variables $\{y_i\}$ and then quantizing them sample by sample. The optimum orthogonal transform is the matrix which most fully exploits the correlations between input samples and minimizes the average mean-square error between input and output.

If the continuous process $x(t)$ is stationary and has an autocorrelation function $R_X(\tau)$, then it is known that in order to achieve the minimum mean integral square error the basis functions should be chosen in the following manner. First, it is necessary to find all eigenvalues and eigenfunctions of the equation

$$\lambda \phi(t) = \int_0^{T_0} R_X(t - \tau) \phi(\tau) d\tau$$

where T_0 is a finite time interval. The eigenvalues are listed in decreasing order, i.e.,

$$\lambda_1 \geq \lambda_2 \geq \lambda_3 \dots$$

The corresponding eigenfunctions are ordered in the same manner, that is, $\phi_1(t)$, $\phi_2(t)$, This makes it possible to represent the Karhunen-Loève expansion of the process $x(t)$ as

$$x(t) = \sum_{i=1}^{\infty} a_i \phi_i \quad t \in [0, T_0]$$

where the coefficients are

$$a_i = \int_0^{T_0} x(t) \phi_i(t) dt$$

Then the mean-square error is

$$\overline{e^2} = E\{[x(t) - \bar{x}(t)]^2\} = R(0) - \sum_{i=1}^{\infty} \lambda_i \phi_i^2(t)$$

The discrete version of this procedure consists of taking N samples of the process and finding the projection of the vector X onto the best orthonormal vectors. The finite Karhunen-Loève expansion for X is

$$X = \sum_{i=1}^N a_i \phi_i \quad (3-38)$$

where ϕ_i , $i = 1, 2, \dots, N$, are the eigenvectors of the covariance matrix* of X :

$$C\phi_i = \lambda_i \phi_i \quad (3-39)$$

*The covariance matrix of X is defined as

$$C = E\{(X - \bar{X})(X - \bar{X})^T\}$$

where \bar{X} denotes the mean value of X and the superscript T denotes transpose.

The eigenvalues λ_i , $i = 1, 2, \dots, N$, are roots of

$$|\lambda_i I - C| = 0 \quad (3-40)$$

where I is an $N \times N$ unity matrix.

The covariance matrix of a real process is symmetric and its eigenvalues are real and positive quantities. The eigenvectors are orthonormal; i.e.,

$$\phi_i^T \phi_j = \delta_{ij} \quad (3-41)$$

Let the ϕ_i 's be ordered so that $\lambda_1 \quad \lambda_2 \quad \lambda_3 \dots \lambda_N$. The transformation matrix K is the transpose of the modal matrix which is formed as

$$K^T = \left\{ \begin{bmatrix} \phi_1 \end{bmatrix} \dots \begin{bmatrix} \phi_N \end{bmatrix} \right\} \quad (3-42)$$

The K^T matrix is the matrix that diagonalizes the covariance matrix C ; i.e.,

$$K^T C K = \Lambda \quad (3-43)$$

where Λ is an $N \times N$ diagonal matrix. The eigenvalues (in decreasing order) appear on the diagonal of Λ . Since C is symmetric and Λ diagonal, equation (3-42) implies that

$$K^{-1} = K \quad (3-44)$$

Thus, if all eigenvalues of the expansion are used, equation (3-37) yields

$$\hat{X} = K^{-1} \hat{Y} = K^{-1} KX = X \quad (3-45)$$

If the expansion of equation (3-38) is truncated to M ($< N$) terms,

$$\hat{X} = \sum_{i=1}^M a_i \phi_i \quad (3-46)$$

Then the mean-square truncation error is

$$\overline{e^2} = E \left\{ (X - \bar{X})^T (X - \bar{X}) \right\} \quad (3-47)$$

Using equations (3-38), (3-41), and (3-46) results in

$$\overline{e^2} = \sum_{i=M+1}^N E \left\{ a_i^2 \right\} \quad (3-48)$$

From the definition of C and from equation (3-39),

$$E \left\{ a_i^2 \right\} = \lambda_i \quad (3-49)$$

Thus, the result is

$$\overline{e^2} = \sum_{i=M+1}^N \lambda_i \quad (3-50)$$

Equation (3-50) gives the minimum error committed when M terms are used in the expansion. Since the λ_i are arranged in decreasing order, the truncation error decreases as more and more terms are added to the expansion.

Implementation of this transformation follows the method outlined previously. The eigenvalues and eigenvectors of the covariance matrix are computed according to well-known numerical techniques.^{31, 32}

4. ORTHOGONAL TRANSFORM PROCESSING OF SPEECH

4.1 GENERAL

The ability of discrete forms of the Fourier, Hadamard, Haar, and Karhunen-Loève transforms to reduce the bit rate necessary to transmit speech signals is examined in this section. To rate the effectiveness of each transformation method in accomplishing this goal, the resulting quantizing error (or noise) in each method is determined at various bit rates and compared with that of conventional companded PCM processing. Computed and experimentally measured results, as well as results obtained from PB word list testing, are presented and compared. This comparison shows that the bit rate required for companded PCM is reduced by 13.5 kbps by using the Karhunen-Loève transform, 10 kbps by using the Fourier transform, 7.5 kbps by using the Hadamard transform, and 7.0 kbps by using the Haar transform. These bit rate reductions are found to be somewhat independent of the transmission bit rate.

4.2 INFORMATION RATE REDUCTION USING ORTHOGONAL TRANSFORMS

A general expression for a discrete time series, $\chi(nT)$, possessing a total signal power σ^2 , in terms of a set of orthogonal coefficients $\{C_i\}$ and orthogonal functions $\{\phi_i(nT)\}$, is given by the transformation

$$\chi(nT) = \sum_{i=0}^{N-1} C_i \phi_i(nT) \quad (4-1)$$

In this expression, T is the time between samples, n is the sample index and N is the number of samples per processing window. The window size, NT , is not related to any specific feature of the signal, such as the pitch period of the formant structure of speech. Instead, it is based on the effective spectral resolution needed to represent the characteristics of the speech spectrum for an ensemble of talkers and words. (The choice of window size is discussed in greater detail at the end of Section 4.3.) The set of orthogonal functions may be any one of many different types. The types of transforms used are the discrete forms of the Fourier, Hadamard (Walsh), Haar, and Karhunen-Loève transforms.

The present objective is to determine a general expression for the information rate needed to transmit the coefficients so that the time series may be reconstructed with the same signal-to-quantizing noise ratio achieved through conventional PCM transmission. This can be done in terms of the set of variances $\{\sigma_i^2\}$ which are in one-to-one correspondence with the set of coefficients $\{C_i\}$. Assume that coefficient C_i with variance σ_i^2 is quantized by using n_i binary digits. Then the quantizing error or noise associated with coded transmission of this coefficient is

$$\overline{\epsilon_i^2} = \frac{K_i}{N^2} \frac{\sigma_i^2}{2^{2n_i}} \quad (4-2)$$

where K_i is a constant that expresses the consequence of the scaling and companding associated with quantization and N^2 is a normalizing factor. N is the number of samples in the time series set involved in determining the coefficients, as well as the number of orthogonal coefficients required to represent a set

of time series samples of size N . The total quantizing noise contributed by quantized transmission of orthogonal coefficients is then

$$\overline{\epsilon_Q^2} = \frac{1}{N^2} \sum_{i=0}^{N-1} K_i \frac{\sigma_i^2}{2^{2n_i}} \quad (4-3)$$

If the same information were transmitted by using PCM, the quantizing error or noise would be

$$\overline{\epsilon_{PCM}^2} = K_{PCM} \frac{\sigma^2}{2^{2M}} \quad (4-4)$$

where M is the total number of bits used to quantize each sample, σ^2 is the signal power (which is also the signal variance, since it is assumed that bias is removed), and K_{PCM} is a constant that expresses the consequence of scaling and companding.

The quantizing noise expressed by equation (4-3) is minimized when the contributions are all equal. This can be shown by equating all partial derivatives with respect to the variables $\{n_i\}$ to zero in equation (4-3), and observing the constraint that $\sum_{i=0}^{N-1} n_i = NM$, where NM is the total number of bits assigned to each window. Thus, if the sum of N contributions is to equal the PCM quantizing noise, each contribution must equal $\overline{\epsilon_{PCM}^2}/N$ and the following equality must be satisfied for each coefficient:

$$\overline{\epsilon_i^2} = \frac{\overline{\epsilon_{PCM}^2}}{N} \quad (4-5)$$

Using equations (4-3) and (4-4) in the preceding equation results in the following expression for the number of bits required for each coefficient to fulfill the condition of quantization noise equality:

$$n_i = M + \frac{1}{2} \log_2 N \frac{K_i}{K_{PCM}} + \log_2 \frac{\sigma_i}{N\sigma} \quad (4-6)$$

The total bit rate for the transmission of the quantized and coded coefficients is thus

$$R_o = \frac{1}{NT} \sum_{i=0}^{N-1} n_i, \text{ bps} \quad (4-7)$$

where T is the time between speech samples. The bit rate for transmitting the same signal with the same quantizing noise by conventional PCM is

$$R_{PCM} = \frac{M}{T}, \text{ bps} \quad (4-8)$$

If the orthogonal representation is to reduce the bit rate, the average bit assignment must be less than M. Thus,

$$\frac{1}{N} \sum_{i=0}^{N-1} n_i < M \quad (4-9)$$

The results given in equations (4-3) and (4-4) can also be used to derive an expression for the improvement in the signal-to-quantizing noise ratio over that achieved with PCM

processing. This improvement will be designated as Λ_Q . Since the signal power will be the same for the orthogonal transform processed signal and the PCM processed signal, the improvement is simply expressed as the logarithm of the ratio of the quantizing noises given by equations (4-3) and (4-4), respectively. That is,

$$\begin{aligned}\Lambda_Q &= 10 \log \left(\frac{\overline{\epsilon_{PCM}^2}}{\overline{\epsilon_O^2}} \right) \\ &= 10 \log_{10} \left(\frac{N^2}{2^{2M} \sum_{i=0}^{N-1} \frac{K_i}{K_{PCM}} \frac{\sigma_i^2}{\sigma^2} 2^{-2n_i}} \right) \quad (4-10)\end{aligned}$$

The values of n_i are, of course, chosen to minimize the quantizing noise resulting from the transform processing. To calculate the values of Λ_Q given in the following sections, it is assumed that $K_i = K_{PCM}$, since the companding law used for all of the orthogonal coefficients is the same as that used for the PCM processing.

To estimate the value of Λ_Q for the speech which was remade after being processed by the various transformations and bit rates, the signal-to-quantizing noise ratios for PCM were computed at the 8000-Hz sampling rate with two, four, and seven bits per sample. The resulting data rates were 16, 32, and 56 kbps using $\mu = 100$ companding and a peak-to-peak quantizing range of 10σ . The computed values of S/N_Q were 2.98, 15.1, and 33.3 dB, respectively. However, in the transformation analysis performed during this study, the transmission rates were 14, 28, and 56 kbps. Therefore, to compare PCM processing at these rates, it was necessary to interpolate the S/N_Q values. This resulted in S/N_Q values of 2.9, 12.1, and 33.3 dB for PCM data

rates of 14, 28, and 56 kbps, respectively. Estimated values of Λ_Q , designated as $\hat{\Lambda}_Q$, were derived by taking the difference between the preceding S/N_Q values for PCM and the S/N_Q values obtained by using particular combinations of transform technique and bit rate.

The spectral distribution of the quantizing noise generated when coefficients are quantized by using the bit assignments given by equation (4-6) is very nearly white. This follows from the spectral distributions of the orthogonal functions and the adjustment of the number of quantizing levels to produce an equal value of quantizing error for all of the orthogonal functions.

4.3 SPEECH SAMPLING AND ANALYSIS OF WINDOW SIZE

To arrange for processing, the speech is sampled at a rate of 8000 samples per second and linearly quantized using 14 bits per sample, i.e., one bit for sign and 13 bits for magnitude. Sampling at a rate of 8000 samples per second preserves the spectral components of the speech signal up to frequencies of 3600 Hz without excessive interference due to aliasing. This is the reason for transmitting all telephone system PCM speech at 8000 samples per second. The use of 14 bits per sample provides an excellent signal-to-quantizing noise ratio (74 dB for speech having a peak-to-rms ratio of 15 dB) and ensures that the quantizing noise contribution caused by the initial data preparation is rather inconsequential as compared to that introduced by the processing methods under investigation.

All processing was performed on an IBM 360/65 computer. The speech data was converted from analog to digital form and vice-versa by using an IBM 1827 data link to the IBM 360/65.

The size of the analysis window is determined by the number of signal samples, N , entering into the computation of the orthogonal coefficients. If these samples are taken at intervals of T seconds, then the analysis window has a time duration of NT . For Fourier analysis, it is well known that the frequency resolution of this window is given by $(NT)^{-1}$. If the spectral distribution property of the signal is to be responsible for the differences in the coefficient variances that result in the information rate reduction, then the resolution provided by the window must be sufficient to resolve these differences.

In the analyses performed during this program, N has been assigned a value of 16. One reason for selecting this value was that it is an integral power of 2; this permits the use of the fast processing form of the orthogonal transformations. Also, this value provides a frequency resolution of 500 Hz for $T = 125$ microseconds, which is sufficient to resolve the long-term spectral distribution of speech in terms of the variances of the Fourier coefficients.

The value of $N = 64$, which provides a frequency resolution of 125 Hz, has also been examined and has been found to improve the efficiency of the representation slightly. Higher values of N will provide greater resolution capacity, easily capable of resolving the short-term formant spectrum structure of speech; however, use of this short-term spectrum structure to reduce information rate requires frequent recomputation of coefficient variance and reassignment of quantizing levels. This latter form of orthogonal signal processing, referred to as adaptive signal processing, was not investigated in great detail.

The rationale used in the previous paragraphs to determine the sample window size for the Fourier transformation also applies approximately to the Hadamard and Karhunen-Loève

transformations, since the spectral distributions of the orthogonal functions of these latter transforms are very similar to those of the Fourier set.

4.4 FOURIER TRANSFORM PROCESSING OF SPEECH

To perform an experimental analysis of the Fourier processing method, a speech sample of 65-second duration was subjected to processing. This same sample was also used as the source data for the other orthogonal signal processing methods tested and discussed in the following paragraphs, and for the PCM processing, which was used as a basis for comparing the effectiveness of each method. The 65-second speech sample comprised the speech of four talkers, two male and two female, each speaking the same five sentences. The sentences used are listed in Table 4-1.

Table 4-1. Test Sentences Used
for Experimental Evaluations

- | |
|--|
| <ol style="list-style-type: none">1. Alibaba knew how men buy lime.2. She caught Sue's fuscia scarf.3. Father let Vee shout four.4. High altitude jets fly by screaming.5. See the old pig's azure foot. |
|--|

For this processing, a window size of $N = 16$ was employed. Since the speech signal was sampled at 125-microsecond time intervals, the window width was two milliseconds. This resulted in an $x^{-1} \sin x$ spectrum window, shown in Figure 4-1, for each Fourier coefficient. The time wave shapes of each of the 16 eigenfunctions are shown in column 1 of Figure 4-2. It should be noted that the eigenfunctions are, of course, discrete.

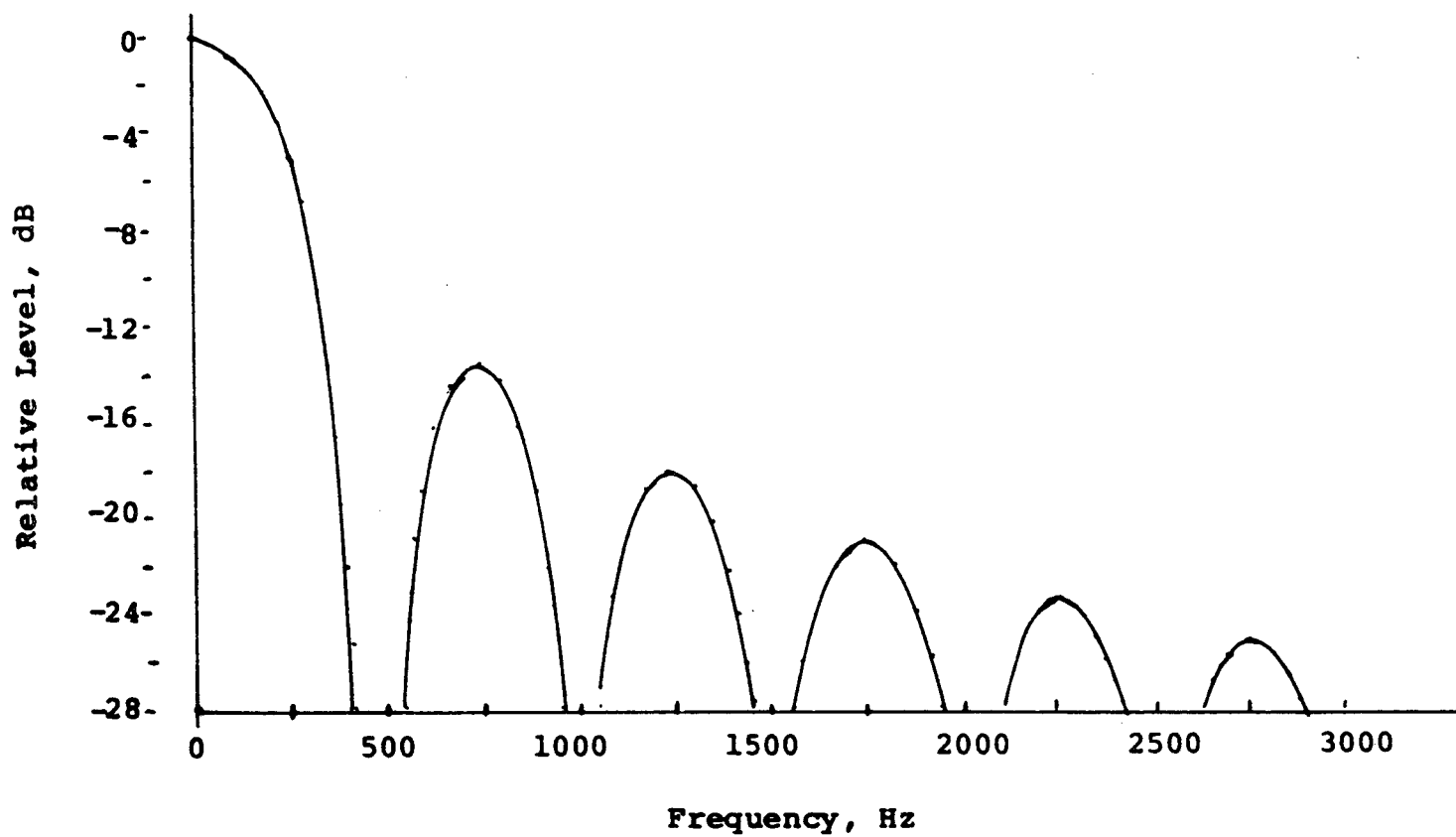


Figure 4-1. Spectral Distribution of the Fourier Coefficient $i = 0$

2

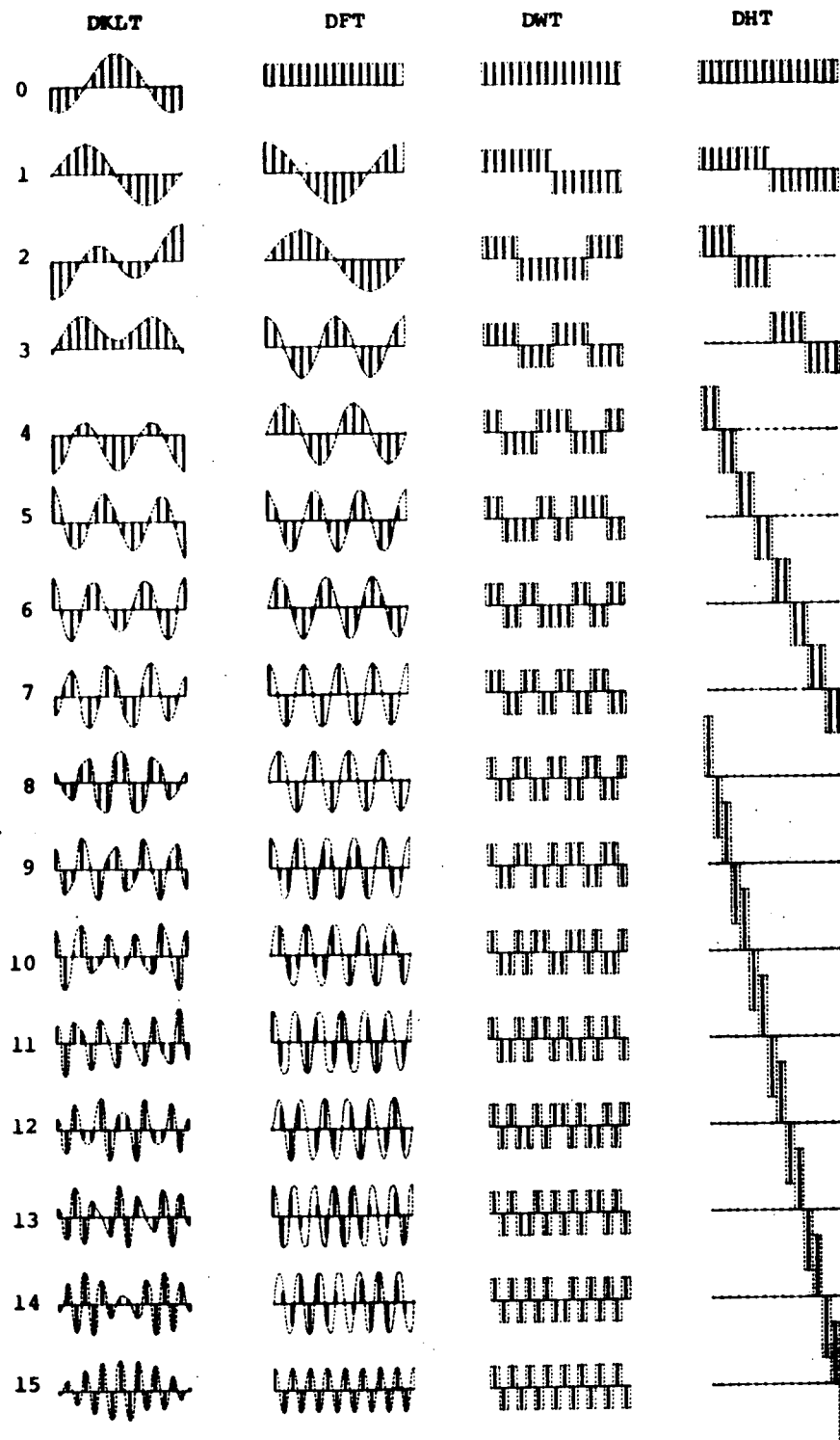


Figure 4-2. Discrete Orthogonal Functions

Quantization of the orthogonal coefficients in the transform domain was performed in a nonlinear fashion. The form of the nonlinearity was chosen to be logarithmic and specifically the same as the PCM companding law generally used for speech. The rationale behind this choice was the same as that for speech. It was an attempt to sacrifice performance at high signal levels to achieve improved performance at low signal levels. The net result is to extend the dynamic range of the system so that a wide range of talker levels can be accommodated equally well.

The companding law employed was

$$y(x) = x_{\max} \frac{\log_e \left(1 + \mu \frac{x}{x_{\max}} \right)}{\log_e (1 + \mu)} \quad \text{if } 0 \leq x \leq x_{\max}$$

and

$$y(x) = -x_{\max} \frac{\log_e \left(1 - \mu \frac{x}{x_{\max}} \right)}{\log_e (1 + \mu)} \quad \text{if } -x_{\max} \leq x \leq 0$$

where x and y are the input and output of the compressor, respectively, and x_{\max} is the peak value of the input which was chosen to be equal to 10σ , where σ^2 is the measured coefficient variance. The value of μ was chosen to be 100. These values were also used for PCM processing of the speech.

Equation (3-14) for the Fourier coefficients was evaluated for the 65-second speech sample by using the FFT method and N equal to 16. The standard deviations of the 16 resulting coefficients were computed next; these are listed in column 1 of Table 4-2. They are designated as σ_i , where i is the coefficient index, and they are, of course, given by the square root of the variance σ_i^2 .

Table 4-2. Fourier Transform ($N = 16$)--The Standard Deviations and the Optimum Bit Assignment for Fourier Coefficients

(improvement $\Lambda_Q = 5.0$ dB, computed from formula)

i	σ_i	n_i (bits)		
		Case 1 ^a	Case 2 ^b	Case 3 ^c
0	5368	8	5	3
1	6156	9	5	4
2	6201	9	5	4
3	3458	8	5	3
4	3979	8	5	3
5	1833	7	4	2
6	2041	7	4	2
7	1334	7	3	1
8	1342	7	3	2
9	1201	7	3	1
10	1021	6	3	1
11	914	6	3	1
12	523	6	2	0
13	857	6	3	1
14	262	5	1	0
15	820	6	2	0

a: $R = 56$ kbps, $S/N_Q = 37.9$ dB, $\hat{\Lambda}_Q = 4.6$ dB

b: $R = 28$ kbps, $S/N_Q = 17.6$ dB, $\hat{\Lambda}_Q = 5.5$ dB

c: $R = 14$ kbps, $S/N_Q = 8.9$ dB, $\hat{\Lambda}_Q = 6.0$ dB

Bit assignments were next made in accordance with the requirement that the rms error, given by equation (4-3), should be minimized. This was accomplished by making n_i , the number of bits assigned to quantizing the i th coefficient, proportional to $\log_2 \sigma_i/\sigma$, where σ (given by the square root of the sum of σ_i^2) is the variance of the total signal. The resulting bit assignments for bit rates of 56, 28, and 14 kbps are shown in Table 4-2.

Based on these bit assignments and the quantizing rule discussed earlier, the 65-second speech sample was processed. The resulting rms error, designated as N_Q , in the processed speech was computed by taking differences, in one-to-one correspondence, between the input and output speech samples. The resulting speech signal-to-quantizing noise ratios, designated as S/N_Q for each of the bit rates, are given in Table 4-2.

Next, the S/N_Q was determined for the same speech sample processed by using conventional PCM techniques and the same companding law. When compared with the S/N_Q values given in the previous paragraphs, the S/N_Q for this speech sample showed an average improvement of 5 dB that was relatively independent of bit rate.

4.5 HADAMARD TRANSFORM PROCESSING OF SPEECH

Hadamard processing was performed by using the same 65-second speech sample which was used for Fourier processing and a window size of $N = 16$. The computation procedure was based on the algorithm given in Section 3.3. The set of orthogonal functions used to represent the speech signal is shown in column 3 of Figure 4-2. Expansion of typical speech wave shapes in terms of the Hadamard set of orthogonal functions exhibited a strong odd harmonic structure that does not occur for the discrete Fourier

set. For this reason, the Hadamard set of orthogonal functions is a less efficient representation of speech than the discrete Fourier set.

The coefficient standard deviations for the 65-second speech sample are given in column 1 of Table 4-2. It can be seen that the nonoptimum representation offered by the Hadamard function set results in values of standard deviation for the higher order coefficients which are higher than those which occur for Fourier processing.

The number of bits per coefficient was adjusted to minimize the mean-squared error for 56-, 28-, and 14-kbps rates by using the same rationale which was used for Fourier processing. The resulting coefficient bit assignments are given in Table 4-3. Also given in this table are the signal-to-quantizing noise ratios calculated by squaring the difference between the processed and unprocessed speech samples. When compared to PCM at 56 kbps, Hadamard processing showed an improvement in signal-to-quantizing noise ratio, Λ_Q , of 3 dB. Hadamard processing shows the same level of improvement at 28 and 14 kbps.

4.6 HAAR TRANSFORM PROCESSING OF SPEECH

Haar processing employs a set of rectangular functions illustrated in column 4 of Figure 4-2 and discussed in Section 3.4. These functions weight the speech samples with decreasing periodicity as illustrated in equation (4-3) for $N = 16$. The Haar coefficients were compared for the 65-second speech sample by using the algorithm given in Section 3.4.2. The resulting standard deviations, which are shown in Table 4-4, are similar to those which result from Hadamard processing in that the values at the higher coefficient indices do not decrease as rapidly as they did in the case of Fourier processing.

Table 4-3. Hadamard Transform ($N = 16$)--The Standard Deviations and the Optimum Bit Assignment for Hadamard Coefficients

(improvement $\Lambda_Q = 3.02$ dB, computed from formula)

i	σ_i	n_i (bits)		
		Case 1 ^a	Case 2 ^b	Case 3 ^c
0	5368	8	5	3
1	7758	9	5	3
2	7543	8	5	3
3	5464	8	4	3
4	4273	8	4	2
5	4459	8	4	3
6	3453	7	4	2
7	2196	7	3	2
8	1526	6	2	1
9	1621	6	3	1
10	1783	6	3	1
11	1787	6	3	1
12	1909	7	3	1
13	2021	7	3	1
14	1553	6	3	1
15	820	5	2	0

a: $R = 56$ kbps, $S/N_Q = 36.1$ dB, $\hat{\Lambda}_Q = 2.8$ dB

b: $R = 28$ kbps, $S/N_Q = 15.6$ dB, $\hat{\Lambda}_Q = 3.5$ dB

c: $R = 14$ kbps, $S/N_Q = 5.4$ dB, $\hat{\Lambda}_Q = 2.5$ dB

Table 4-4. Haar Transform ($N = 16$)--The Standard Deviations and the Optimum Bit Assignment for Haar Coefficients

(improvement $\Lambda_Q = 2.98$ dB, computed from formula)

i	σ_i	n_i (bits)		
		Case 1 ^a	Case 2 ^b	Case 3 ^c
0	526	8	4	3
1	779	9	5	3
2	656	8	5	3
3	653	8	5	3
4	339	8	4	2
5	338	7	4	2
6	343	8	4	2
7	340	8	4	2
8	145	6	3	1
9	145	6	3	1
10	144	6	3	1
11	144	6	3	1
12	144	6	2	1
13	143	6	2	1
14	146	6	3	1
15	143	6	2	1

a: $R = 56$ kbps, $S/N_Q = 35.9$ dB, $\hat{\Lambda}_Q = 2.6$ dB

b: $R = 28$ kbps, $S/N_Q = 14.8$ dB, $\hat{\Lambda}_Q = 2.7$ dB

c: $R = 14$ kbps, $S/N_Q = 5.24$ dB, $\hat{\Lambda}_Q = 2.34$ dB

Coefficient bit assignments were made to minimize the rms error at 56, 28, and 14 kbps. The resulting bit assignments are shown in Table 4-4. Comparison with bit assignments for Hadamard processing shown in Table 4-3 reveals that the Haar coefficients have a slightly more uniform distribution. This indicates that the Haar function set is a slightly less effective representation of the speech process than the Hadamard set.

The resulting calculated signal-to-quantizing noise ratios for each of the bit rates are shown in Table 4-4. These are slightly less than the values achieved for Hadamard and Fourier processing. The improvement in Λ_Q over PCM at 56 kbps was found to be 2.98 dB. These considerations indicate that the processing advantage provided by the Haar transform is approximately the same as that provided by the Hadamard transform.

4.7 KARHUNEN-LOÈVE TRANSFORM PROCESSING OF SPEECH

4.7.1 Introduction

The Karhunen-Loève transform (KLT) was used to process speech at bit rates of 56, 28, and 14 kbps. Results demonstrated better quality at all bit rates than was achieved by using any of the other transform techniques. In terms of signal-to-quantization noise ratio, speech processed by the KLT outperformed all of the other techniques.

The theoretical development of the KLT method was presented in Section 3.5.

4.7.2 Determination of the KLT Matrix

To determine the KLT matrix, it is necessary to obtain the autocorrelation function of the time series that is to be processed. This autocorrelation function makes it possible to form the normalized covariance matrix which is represented by

$$C = \begin{bmatrix} 1 & \rho(1) & \rho(2) & \rho(3) & . & . & . & \rho(15) \\ \rho(1) & 1 & \rho(1) & \rho(2) & & & & \rho(14) \\ \rho(2) & \rho(1) & 1 & \rho(1) & & & & \rho(13) \\ \rho(3) & \rho(2) & \rho(1) & 1 & & & & \rho(12) \\ . & & & & . & & & . \\ . & & & & & . & & . \\ . & & & & & & 1 & \rho(1) \\ \rho(15) & \rho(14) & \rho(13) & \rho(12) & . & . & \rho(1) & 1 \end{bmatrix}$$

where $\rho(k)$ is the autocorrelation coefficient at delay time $k\tau$, and τ is the time between samples. For the processing conducted here, $\tau = 125$ microseconds.

Figure 4-3 gives the autocorrelation functions for the speech samples of the male and female talkers used to demonstrate the method. In each case, the autocorrelation function was obtained by averaging over the entire speech sample for each talker. These speech samples were approximately 15 seconds in duration. It should be noted that 16 autocorrelation values appear in the covariance matrix; this is the number of samples per window to be used in KLT processing. The values to be installed in the covariance matrix are, of course, obtained from the autocorrelation function shown in Figure 4-3.

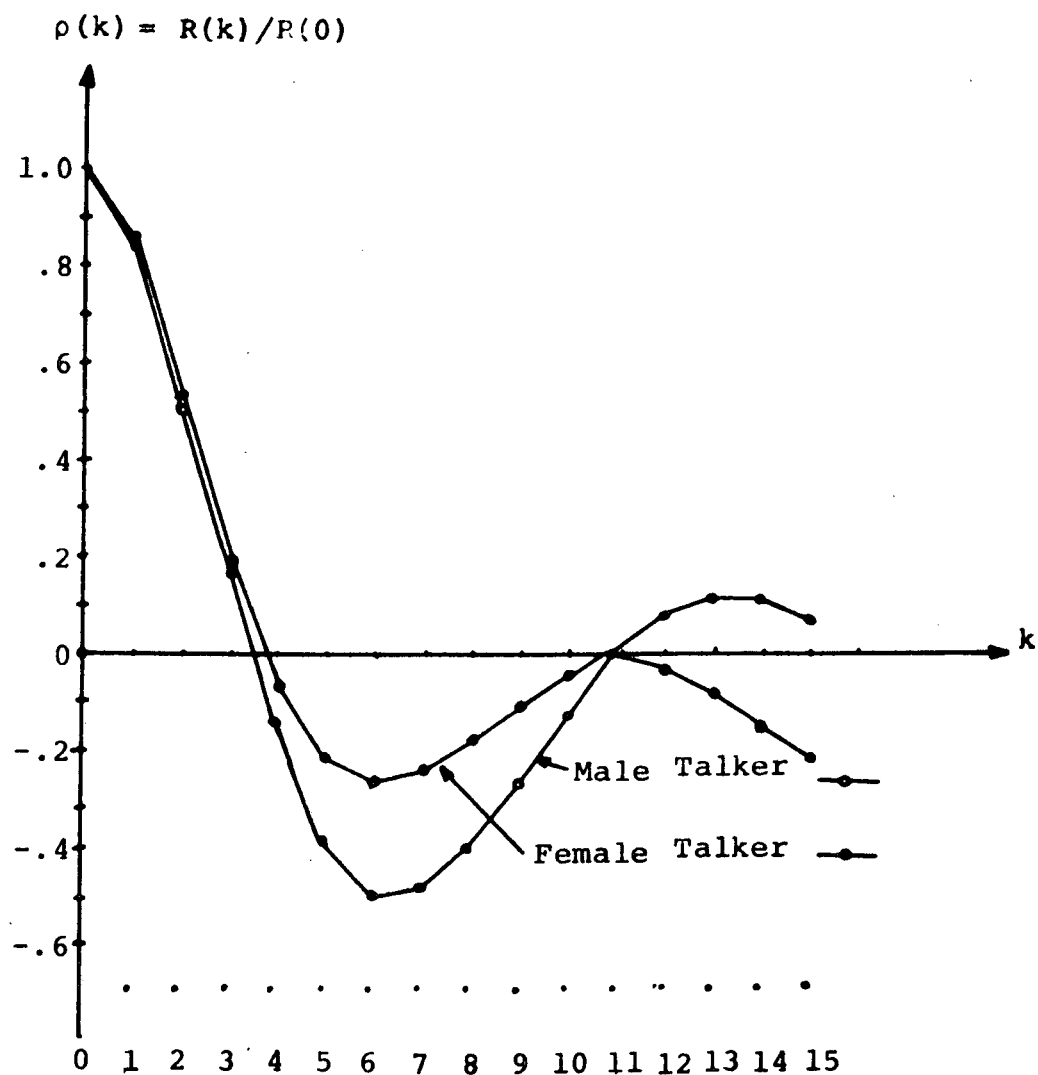


Figure 4-3. Normalized Autocorrelation Functions

The KLT needed to process the speech is the transpose of the matrix that diagonalizes the covariance matrix. The values on the diagonal are the eigenvalues, which are measures of the variances of each of the KLT coefficients. The eigenvalues designated by λ_i are given in Table 4-5. This table also shows the actual values of the variances, designated as σ_i^2 , for each KLT coefficient. These values, which are determined by applying the KLT to the male speech sample, are averaged over the entire 15-second sample using a window size of 16 samples; it is assumed that the same KLT functions apply to each window.

For optimum processing, the ratios of the computed variances to the eigenvalues should be almost a constant. Nevertheless, the ratios which are given in the last column of Table 4-5 vary widely from a constant value. This is because the KLT functions used for computing the coefficients were determined for the entire 15-second speech passage and are not truly optimum for each window. It should be pointed out that, for the speech processing conducted here, even though the same KLT functions were used for all sample windows, extremely good results were obtained. The fact that the variances were not in a constant ratio with the eigenvalues indicates that even further improvement in performance could be obtained by recomputing the KLT functions more frequently during the processing of the speech. However, this is a burdensome process and is considered to be impractical.

Table 4-5. Karhunen-Loève Transform (N = 16)--
Eigenvalues and Computed Variances

i	λ_i	σ_i	$(\sigma_i^2/\lambda_i) \times 10^6$
0	5.402	4270	3.37
1	5.149	4219	3.45
2	2.172	2769	3.52
3	1.627	2324	3.31
4	0.750	1594	3.38
5	0.356	1067	3.19
6	0.156	701	3.15
7	0.104	578	3.21
8	0.098	550	3.08
9	0.092	576	3.60
10	0.051	416	3.39
11	0.022	237	2.55
12	0.012	147	1.80
13	0.007	114	1.85
14	0.001	87	7.56
15	0.0005	38	2.88

The values of the KLT matrix used for computing the coefficients can be arrayed as orthonormal functions; this is done in column 1 of Figure 4-2 for values of the coefficient index, i, ranging from 0 through 15. It can be noted that these functions have many interesting properties. Half of them are even functions and half of them are odd functions. They are orthonormal

in all possible combinations. The DC term is picked up principally in the fourth and fifth coefficients. The functions appear to be sinusoidal or cosinusoidal and are similar to Fourier functions, but they exhibit varying amounts of phase shift and distortion.

4.7.3 Results of Speech Processing Using the KLT

Experiments performed on speech processing using the KLT show that its orthogonal function set represents speech more efficiently than any of the other orthogonal function sets. Achievement of this level of performance enhancement required that a unique set of orthogonal functions be determined for each talker. Thus, one might say that the KLT processing was made talker adaptive (or talker dependent) by virtue of different orthogonal function sets. Such adaptivity is, of course, impossible for any of the other orthogonal function sets since they are specifically determined beforehand.

Based on the criterion of minimum rms quantizing error, the bit assignments used for one talker are given in Table 4-6. For different talkers, the orthogonal function sets and the bit assignments differed slightly. In the 14-kbps case, two bit assignments were explored as shown. The resulting signal-to-quantizing noise ratios are also given in Table 4-6.

The KLT performance at 56 kbps in terms of signal-to-quantization noise ratio exceeds that of conventional PCM by 7.7 dB. This margin increases to 9.8 dB at 28 kbps and to 9.7 dB at 14 kbps. It should also be noted that the performance of the KLT at 14 kbps was equivalent to that achieved by conventional PCM at 28 kbps.

Table 4-6. Karhunen-Loève Transform ($N = 16$)--Optimum
Bit Assignments for KLT Coefficients(improvement $\Lambda_Q = 9.5$ dB, computed from formula)

i	n_i (bits)			
	Case 1 ^a	Case 2 ^b	Case 3 ^c	Case 4 ^d
0	10	6	4	4
1	10	6	4	4
2	9	6	3	4
3	9	6	3	4
4	8	5	3	3
5	8	4	3	3
6	7	4	2	3
7	7	4	2	3
8	7	3	2	0
9	7	3	2	0
10	7	3	0	0
11	6	2	0	0
12	5	2	0	0
13	5	1	0	0
14	4	1	0	0
15	3	0	0	0

a: $R = 56$ kbps, $S/N_Q = 41.0$ dB, $\hat{\Lambda}_Q = 7.7$ dBb: $R = 28$ kbps, $S/N_Q = 21.9$ dB, $\hat{\Lambda}_Q = 9.8$ dBc: $R = 14$ kbps, $S/N_Q = 12.6$ dB, $\hat{\Lambda}_Q = 9.7$ dBd: $R = 14$ kbps, $S/N_Q = 12.3$ dB, $\hat{\Lambda}_Q = 9.4$ dB

The spectrum distribution function of the speech and the quantization noise produced for the 14-kbps operation are shown in Figure 4-4. It is seen that, at the low-frequency portion of the spectrum, the spectral density of the speech exceeds that of the quantization noise by as much as 16 dB, whereas in the high-frequency portion of the spectrum, the spectral density of speech and noise are nearly equal. This distribution function represents the performance of a male talker averaged over a 15-second speech sample.

4.8 PB WORD LIST PROCESSING AND TESTING

The effectiveness of speech communication depends upon the nature of the speech material, the talker and listener, and the conditions under which the transmission and reception occur. A relative measure of the effectiveness of various communication systems can be obtained by standardizing the speech material and by using an intelligibility test procedure. Thus far, the evaluation of various orthogonal transform processing methods has been based on computation of signal-to-quantizing noise ratios. Quality has also been assessed by expert listeners, but this evaluation cannot be expressed in a quantitative form.

To subjectively assess the relative performance of the various processing techniques, PB (phonetically balanced) word list testing has been used. Three word lists (lists 1, 3, and 4 in Table 4-7) have been selected from the 20 different lists (consisting of 50 words each) of the American Standard Method for measurement of monosyllabic word intelligibility. Lists 1 and 3 were recorded by a first talker (SJC), and lists 3 and 4 by a second talker (JAS). The source tape thus contained 200 spoken monosyllabic words. List 3 was read by both talkers to permit assessment of speaker dependency.

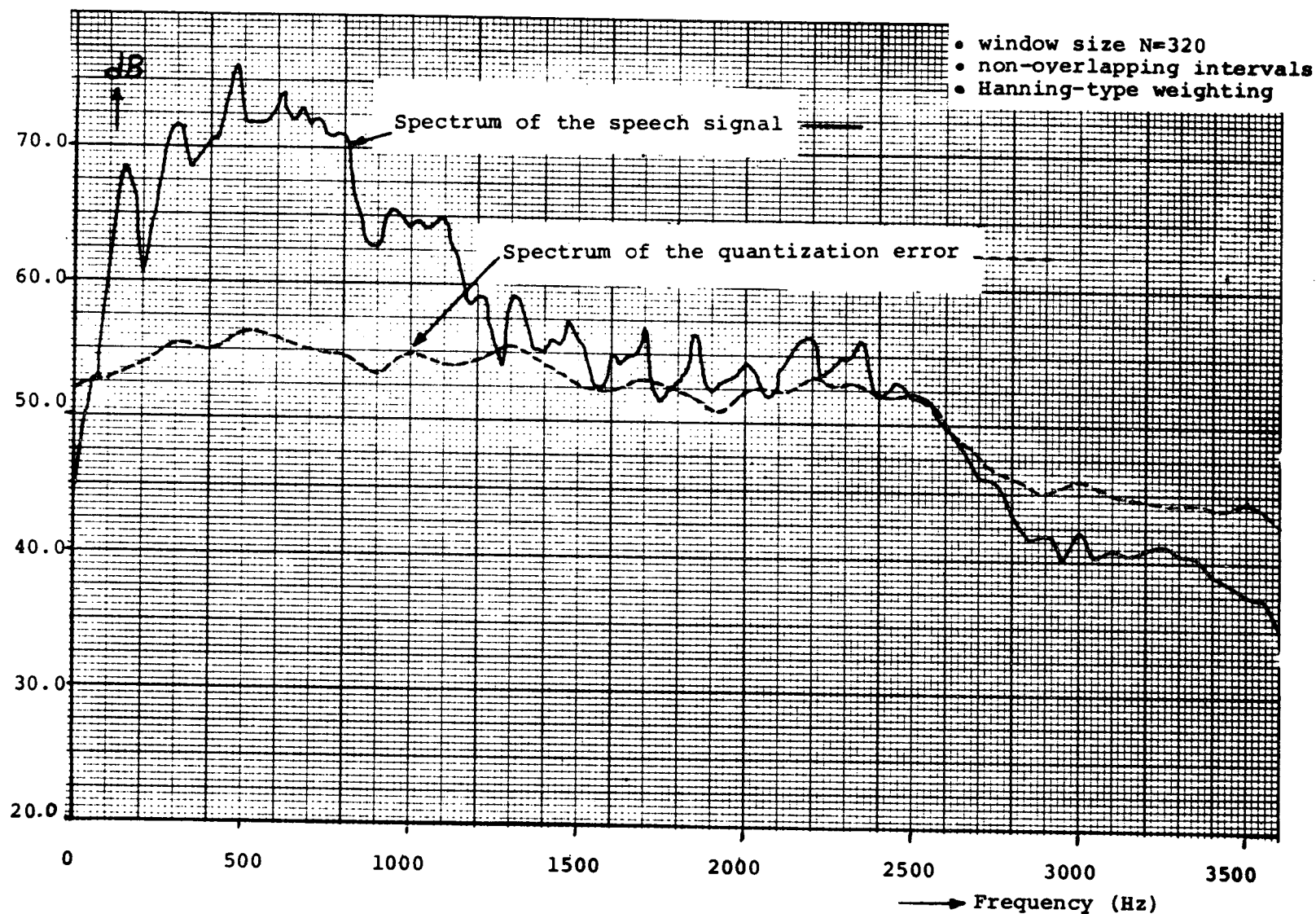


Figure 4-4. Spectral Distribution of Speech Signal and Quantization Error in KLT Processing of 15 Seconds of Speech (case 3 in Table 4-6)

Table 4-7. PB Word Lists

PB Word List No. 1		PB Word List No. 2	
1. cane	26. pest	1. tang	26. blush
2. there	27. slip	2. fate	27. nab
3. dish	28. rub	3. suck	28. bait
4. hid	29. feast	4. else	29. bud
5. heap	30. deed	5. pit	30. rap
6. pants	31. cleanse	6. gill	31. moose
7. hunt	32. folk	7. charge	32. trash
8. no	33. nook	8. bought	33. gloss
9. bar	34. mange	9. cloud	34. perk
10. pan	35. such	10. mute	35. vamp
11. fuss	36. use (yews)	11. bean	36. start
12. creed	37. crash	12. scythe	37. earl
13. box	38. ride	13. vest	38. corpse
14. strife	39. pile	14. rib	39. sludge
15. dike	40. rat	15. pick	40. tan
16. not	41. rag	16. hock	41. ways
17. ford	42. is	17. our	42. bounce
18. end	43. wheat	18. hit	43. niece
19. then	44. rise	19. job	44. awe
20. bask	45. hive	20. wish	45. them
21. fraud	46. grove	21. nut	46. need
22. smile	47. toe	22. dab	47. quart
23. death	48. plush	23. frog	48. five
24. are	49. clove	24. log	49. hire
25. bad	50. fern	25. snuff	50. shoe

Table 4-7. PB Word Lists (Continued)

PB Word List No. 3		PB Word List No. 4	
1. why	26. size	1. float	26. new
2. turf	27. wedge	2. sage	27. rut
3. gnaw	28. deck	3. cloak	28. neat
4. drop	29. hurl	4. race	29. dodge
5. jam	30. wharf	5. tick	30. sketch
6. flush	31. leave	6. touch	31. merge
7. rouse	32. crave	7. hot	32. bath
8. neck	33. vow	8. pod	33. court
9. sob	34. law	9. frown	34. oils
10. trip	35. stag	10. rack	35. shin
11. dill	36. oak	11. bus	36. peck
12. thrash	37. nest	12. blonde	37. beast
13. dig	38. sit	13. pert	38. heed
14. rate	39. crime	14. shed	39. eel
15. far	40. muck	15. kite	40. move
16. check	41. tame	16. raw	41. earn
17. air	42. take	17. hiss	42. budge
18. bead	43. who	18. fin	43. sour
19. sped	44. toil	19. scab	44. rave
20. cast	45. path	20. how	45. bee
21. class	46. pulse	21. strap	46. bush
22. lush	47. fig	22. slap	47. test
23. shout	48. barb	23. pinch	48. hatch
24. bald	49. please	24. or	49. course
25. cape	50. ache	25. starve	50. dupe

The source tape was digitized at the sampling rate of $f_s = 8000$ Hz. An interval of three to four seconds between words was used to permit a listener to write down his response. In order to reduce the computer processing time, this tape was processed by using a voice switch to eliminate processing during the silence periods. Figure 4-5 shows the computed autocorrelation functions of the two talkers.

Fourier, Hadamard, and Karhunen-Loève transformations and companded PCM were used to process the PB word lists. For each transformation type and talker, a set of coefficient variances was computed. These were then used to assign the number of quantizing bits so that the quantizing error power was uniform as discussed in the previous section. For bit values of 56, 28, and 14 kbps, the bit assignments for each of the coefficient sets are given in Table 4-8. Note that, for the Karhunen-Loève transformation, the bit assignments differ for each talker. It should be recalled that, for this transformation, the orthogonal functions are also different for each talker. Attempts to process the speech of two talkers by using a compromise set of orthogonal functions (determined from the autocorrelation function of a composite of the speech of both talkers) have led to results which are inferior to those achieved when the orthogonal functions specifically determined for each talker are used.

Table 4-9 lists the speech signal-to-quantizing noise ratios for each of the orthogonal processing techniques for each talker. As was the case for the spoken sentences used for the analysis given in the preceding section, the methods in order of decreasing signal-to-quantizing noise ratio are: Karhunen-Loève, Fourier, and Hadamard. In addition, it should be noted that the values of signal-to-noise ratio achieved are less than those obtained for the sentences. This is probably attributable to the

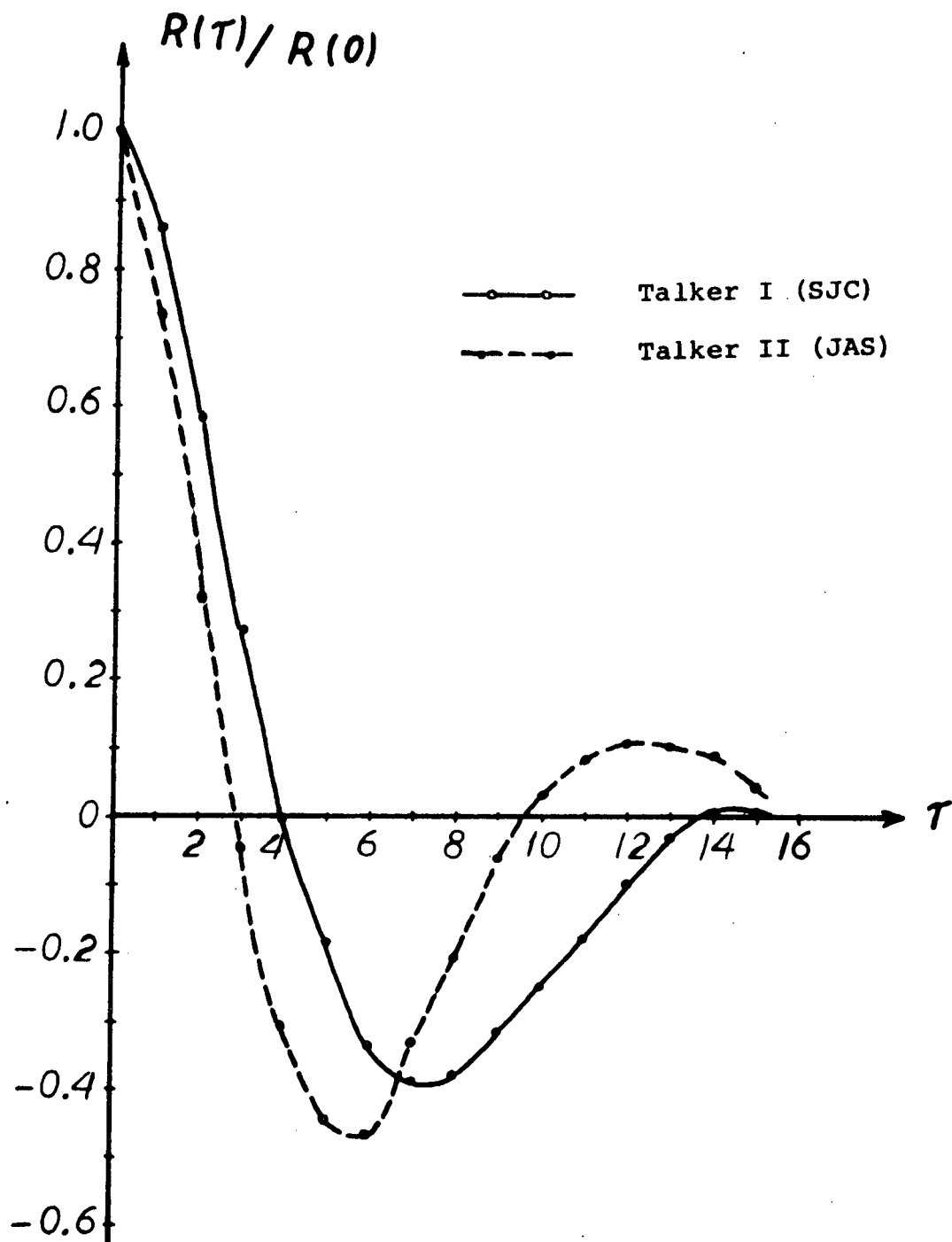


Figure 4-5. Autocorrelation Functions, Based on 100 PB Words, for Two Talkers

Table 4-8. Coefficient Bit Assignments for PB Word List Processing

Coefficient Index	Fourier			Hadamard			Karhunen-Loève					
	Bit Rate (kbps)			Bit Rate (kbps)			Speaker SJC Bit Rate (kbps)			Speaker JAS Bit Rate (kbps)		
	56	28	14	56	28	14	56	28	14	56	28	14
	Coefficient Bit Assignment											
0	9	6	4	9	6	4	9	6	4	9	5	4
1	9	5	3	9	5	3	9	6	4	9	5	4
2	9	5	3	9	5	3	9	5	3	8	5	3
3	8	5	3	8	4	3	9	5	3	8	5	3
4	8	5	3	8	4	3	8	5	3	8	4	2
5	7	5	2	7	4	2	8	4	2	7	4	2
6	7	5	2	7	4	2	7	4	2	7	4	2
7	7	4	2	7	3	2	7	3	2	7	4	2
8	7	4	2	7	3	2	6	3	1	7	3	2
9	6	4	2	6	3	2	6	3	1	7	3	1
10	6	4	2	6	3	2	6	2	1	6	3	1
11	6	2	0	6	3	1	6	3	1	6	3	1
12	6	2	0	6	3	1	6	2	1	6	3	1
13	6	0	0	6	3	1	6	2	0	6	2	0
14	6	0	0	6	3	1	6	2	0	6	2	0
15	5	0	0	5	0	0	4	1	0	5	1	0

noncontinuous, low-duty cycle character of the PB words as opposed to the continuously flowing, high-duty cycle character of the sentences.

Table 4-9. Signal-to-Quantizing Noise Ratios for PB Words

Bit Rate (kbps)	Signal-to-Quantizing Noise Ratio							
	Karhunen-Loève		Fourier		Hadamard		PCM*	
	SJC	JAS	SJC	JAS	SJC	JAS	SJC	JAS
56	40.1	37.4	36.1	34.8	34.75	33.4	32.6	32.7
28	19.9	16.8	16.5	14.4	15.8	13.9	12.1	12.1
14	9.8	7.8	7.5	6.7	6.5	5.5	3.0	3.0

*28- and 14-kbps values are interpolated from the values computed for 56, 32, and 16 kbps.

The PB words processed as described previously were scored by a jury of four experienced listeners; the results are tabulated in terms of percentage of words correctly perceived as a function of bit rate and processing method in Table 4-9.

The tape recordings used for these tests were given to NASA/MSC for more intensive evaluation, but the results were not available at the time that this report was prepared.

4.9 ADDITIONAL ATTEMPTS TO REDUCE BIT RATE

In this section, two methods to further reduce the bit rate needed to transmit the orthogonal transform coefficients are discussed. The first of these, in which the quantizing noise spectrum was adjusted to more nearly follow the shape expected

for typical speech, appears to have some merit. The second, in which the orthogonal coefficients of the Fourier transform were coded in a polar form rather than in a Cartesian coordinate form, has not produced significant data rate reduction.

4.9.1 Variation of Coefficient Bit Assignments

Judicious assignment of the number of quantizing levels allowed for each orthogonal coefficient may make it possible to further reduce the bit rate for intelligible speech. The cause of this reduction is the average spectrum distribution function of speech and particularly the manner with which the speech spectrum rolls off at high frequencies. Data rate reduction may be achieved by assuming an average speech spectrum distribution, determining the relationship for the noise spectrum as a function of various contiguous bands in the spectrum for both PCM and coded Fourier cases, and assigning quantization levels in the coded Fourier case so that the quantization noise in each reconstructed band is proportional to the speech spectrum density. When this is done, the speech spectrum is uniformly masked by the quantizing noise and some further data rate reduction is possible.

Several bit assignment arrangements stemming from the preceding approach were tested for the fast Fourier and fast Hadamard transforms using a 16-sample window size. The arrangements were evaluated by listener assessment. The bit assignments used are listed in Table 4-10. Each arrangement tested is identified by a letter of the alphabet, and *i* indicates the order of the coefficient.

Table 4-10. Assignments of the Number of Bits per Orthogonal Coefficient for Various Speech Processing Schemes

Coefficient for various arrays																		
Array	Bit Rate (kbps)	i (Order of the Coefficient)																Total Bits
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
		Bits per Coefficient																
A	49.5	8	8	8	7	7	7	7	6	6	5	5	5	5	5	5	5	99
B	38.0	0	8	8	7	7	7	7	6	6	5	5	5	5	0	0	0	76
C	36.0	0	6	6	6	6	6	6	6	6	6	6	6	6	0	0	0	72
D	42.0	8	8	8	7	7	7	7	6	6	5	5	5	5	0	0	0	84
E	32.5	5	5	5	5	5	5	5	5	5	5	5	5	5	0	0	0	65

These bit assignment arrangements produced higher values of computed rms error, but, because the error distribution as a function of frequency either falls in an unused part of the spectrum (such as above 3200 Hz or below 300 Hz) or is distributed so that it is less disturbing to the human hearing mechanism (i.e., shaped in a manner that uniformly distributes articulation index loss across the speech spectrum), the loss in intelligibility at a given value of rms error is reduced.

Arrangements B, C, D, and E in Table 4-10 are attempts to obtain acceptable quality and intelligibility at low data rates. The rationale behind arrangements B and C is that the speech signal might be effectively bandpassed from 250 Hz to 3.25 kHz by not transmitting the $i = 0, 13, 14, 15$ coefficients. The remaining coefficients are either assigned the same bit values as the coefficients of arrangement A or are all given a value of six bits. Consequently, it should be expected that energy in the ranges from 0 to 250 Hz and from 3250 to 4000 Hz is removed from the final signal. Of course, the spectrum of the removed coefficients extends into the range above 250 Hz and below

3250 Hz, and the spectrum of the remaining coefficients spreads into the range below 250 Hz and above 3250 Hz, since the spectrum shape of the coefficients is of the form:

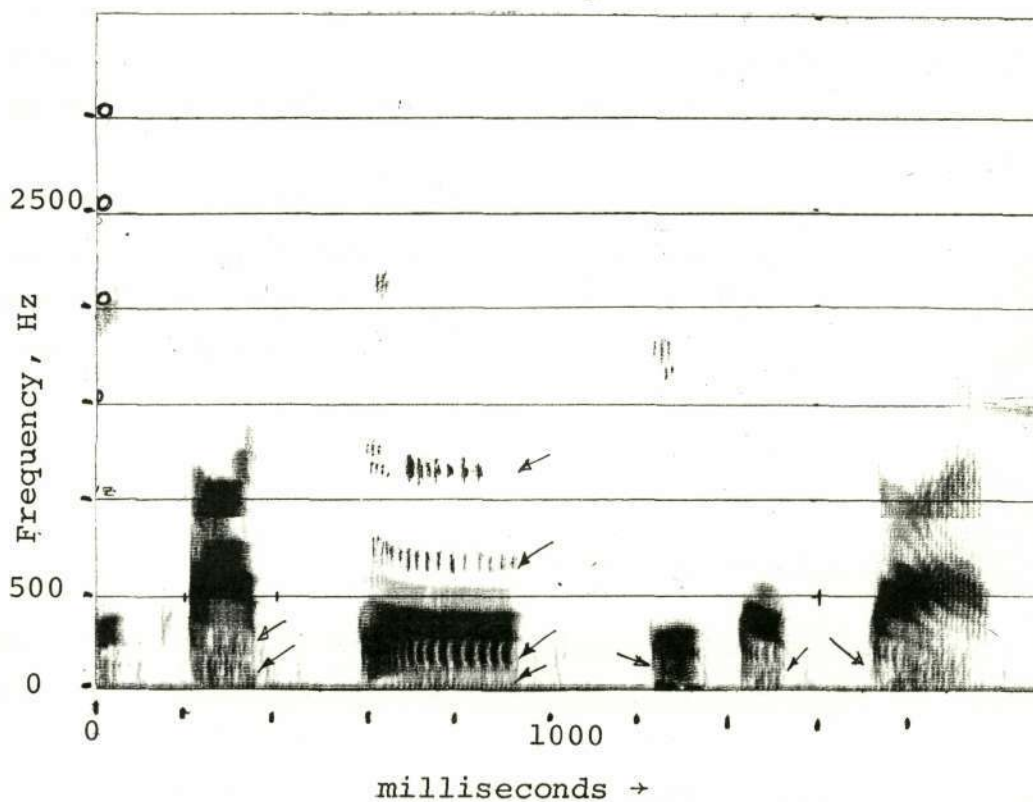
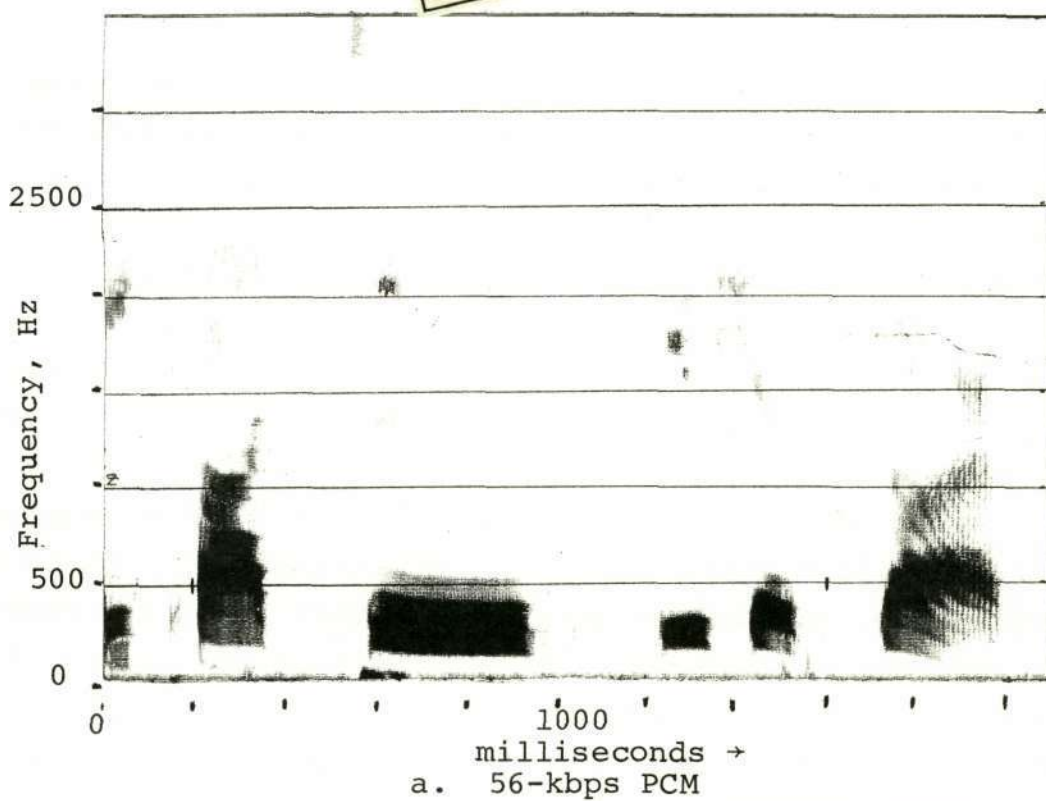
$$S(\omega_i) = \sigma_i^2 T^2 \left[\frac{\sin (\omega - \omega_i) (T/2)}{(\omega - \omega_i) (T/2)} \right]^2$$

where ω_i is the frequency of the i th Fourier coefficient and σ_i^2 is its variance. This spectrum shape was plotted in Figure 4-1 for the window size $T = 2$ ms, which is the window width for 16 samples taken at 8000 samples per second.

Figures 4-6a and 4-6b are sonagrams which clearly show the result of dropping the $i = 0$ coefficient. Figure 4-6a is for conventional 56-kbps speech, and Figure 4-6b is for speech produced by bit arrangement B. The arrows indicate areas where the undesired effects of dropping the $i = 0$ coefficient are present. Undesired effects occurring below 250 Hz are, of course, removed when the speech is high-passed at 250 Hz, but other effects seen above 250 Hz cannot be removed so simply. It is obvious that the effect of dropping a coefficient does extend into the space occupied by adjacent coefficients and, hence, some audible effect should be expected.

Arrangements B and C were designed to explore the effect of dropping the lowest and highest order coefficients. Dropping the 0-order coefficient was found to produce a raucous, husky effect that is considered to be unacceptable. However, dropping the three highest order terms produced a result that is acceptable but which occasionally results in a slight "ringing" effect for female voices.

As previously pointed out, arrangements B and C also differed in the manner of assigning the bits; B was assigned values that minimized the computed mean-squared error and C was

Figure 4-6. Effect of Removal of $i = 0$ Coefficient

assigned values that produced a constant signal-to-noise ratio in each coefficient spectrum range. The latter was designed to produce a uniform loss in terms of articulation index over the speech spectrum range. For this arrangement, the assignment was six bits per coefficient, producing a signal-to-quantizing noise ratio of approximately 30 dB in each coefficient spectrum range. There was very little difference between the results of arrangements B and C. Hence, it was decided that the technique of bit assignment equalization has merit.

Arrangements D and E were designed to further explore the merit of uniform bit assignment. Arrangement D which is also a reference case, is the same as A except that the $i = 0$ coefficient is restored so that the raucous effect caused by its absence is eliminated. Arrangement E, which also restores the $i = 0$ coefficient, uses a constant bit assignment of five bits for all of the coefficients except the top three. Both of the D and E arrangements produced good results. The latter is particularly interesting, since it operates at the relatively low bit rate of 32.5 kbps.

All arrangements listed in Table 4-10 were tested by Fourier processing. Arrangement E was tested by both Fourier and Hadamard processing. The result of Hadamard processing demonstrated very good quality and intelligibility and even seemed to sound better than the results of Fourier processing.

4.9.2 Polar Coordinate Representation

Another method of transmitting the Fourier orthogonal coefficients is made possible by expressing the real and imaginary parts of each coefficient in terms of a polar coordinate representation, i.e., a resultant or magnitude and a phase angle.

It was hoped that this representation would permit some reduction in the bit rate required to transmit the magnitude information, since the magnitudes are related to the formant structure of the speech spectrum and should therefore change at a slow rate. The investigation, however, showed that, because of the arbitrary phase relationship between the periodicities contained in the speech signal and the sampling periodicity, the change in magnitudes is sufficiently rapid to discount the rate reduction that had been desired. It is likely that this difficulty could have been avoided by significantly increasing the window size, e.g., to $N = 256$, so that averaging sufficient to remove the periodicity interaction would occur; however, this would significantly increase the implementation complexity and was considered impractical at this time.

The polar coordinate technique was first attempted for windows of sample size 16. The magnitudes were averaged for sequences of eight windows and coded in companded quantized form with five bits. The phases, of course, were not averaged; a full set of phase values was transmitted for each window. The bit rate that would be produced for this kind of processing is 26 kbps. However, in an initial analysis of this processing method, it was found that the magnitudes of the Fourier coefficients computed for sets of 16 samples varied significantly from window to window. Hence, the average value taken over a set of eight windows would not accurately represent the values of the resultants, since a window size of 16 samples is too short a time interval (only 2 milliseconds) to contain a sample of the speech that is representative of the long-term average. That is, the phase of the window relative to the phase of the fundamental pitch of the talker's speech will cause significant variations in the values of the resultants. Hence, an average value taken over a set of, for example, eight windows will not accurately represent the correct values which occur for each window.

One possible means of circumventing this problem is to increase the window duration by increasing the number of samples per window. To accomplish this, it was decided to use a window width containing 64 speech samples and to average over eight windows. This strategy would result in a bit rate of 26 kbps. This window would contain a segment of speech that is less likely to change significantly as the phase of the window changes relative to the phase of the fundamental pitch of the speech. With five bits per coefficient to represent the average resultants and six bits per phase coefficient, this method was used to process speech. Analysis of the values of the resultants from window to window showed that some variations still existed even with the longer window.

The method described in the preceding paragraph was used to process speech by the computer. The remade speech possessed a rough quality that was judged unsatisfactory. This quality can only be attributed to the failure of the averaging process to accurately represent the true values of the resultants when averaging over eight windows.

It was next decided to average over only two windows using the 64-sample size window. This produced a speech quality which was considerably improved over that achieved by averaging over eight windows, but was still unsatisfactory. The two-window averaging required a bit rate of 34 kbps. However, the reduction in bit rate is probably not worth the quality reduction that results.

In summary, it appears that the polar coordinate representation in itself, when no averaging is applied, produces very good speech quality. When averaging was applied in an attempt to reduce bit rate, the reduced quality which was associated with the bit-rate reduction more than offset any value of the reduction. Some consideration was also given to further increasing

window size; however, it was felt that any further increase in window size would not permit averaging to be performed at all, since the speech spectrum would have changed significantly during the time lapse from one window to the next. Therefore, it was concluded that averaging the magnitudes of the polar coordinate representation over a number of windows is not a good method for further reducing the bit rate for orthogonal speech transmission.

4.10 SUMMARY OF THE ORTHOGONAL TRANSFORM TECHNIQUES FOR SPEECH PROCESSING

Four orthogonal transformations, Fourier, Hadamard, Haar, and Karhunen-Loève, have been evaluated for speech processing and compared in terms of signal-to-quantizing noise performance and PB word list scores. In each case, the same 65-second duration speech sample, containing the speech of two male and two female talkers, was used to perform the experimental analysis. The source speech was sampled at a rate of 8000 samples per second, and processing was performed using window sizes of $N = 16$. Speech was processed by each of the transforms at bit rates of 56, 28, and 14 kbps. In addition to orthogonal transform processing, PCM processing, optimally adjusted to produce the best possible performance, was employed. This procedure constituted a basis for ranking the performance of the orthogonal processing techniques.

The performance of each of the processing methods in terms of signal-to-quantizing noise ratio as a function of bit rate is shown in Table 4-11 and Figure 4-7. Also shown in Table 4-11 are the signal-to-quantizing noise ratios for PCM processing (interpolated in the case of 28 and 14 kbps from the nearest values which are integral multiples of 8000 samples/second). In addition, for each processing method and bit rate,

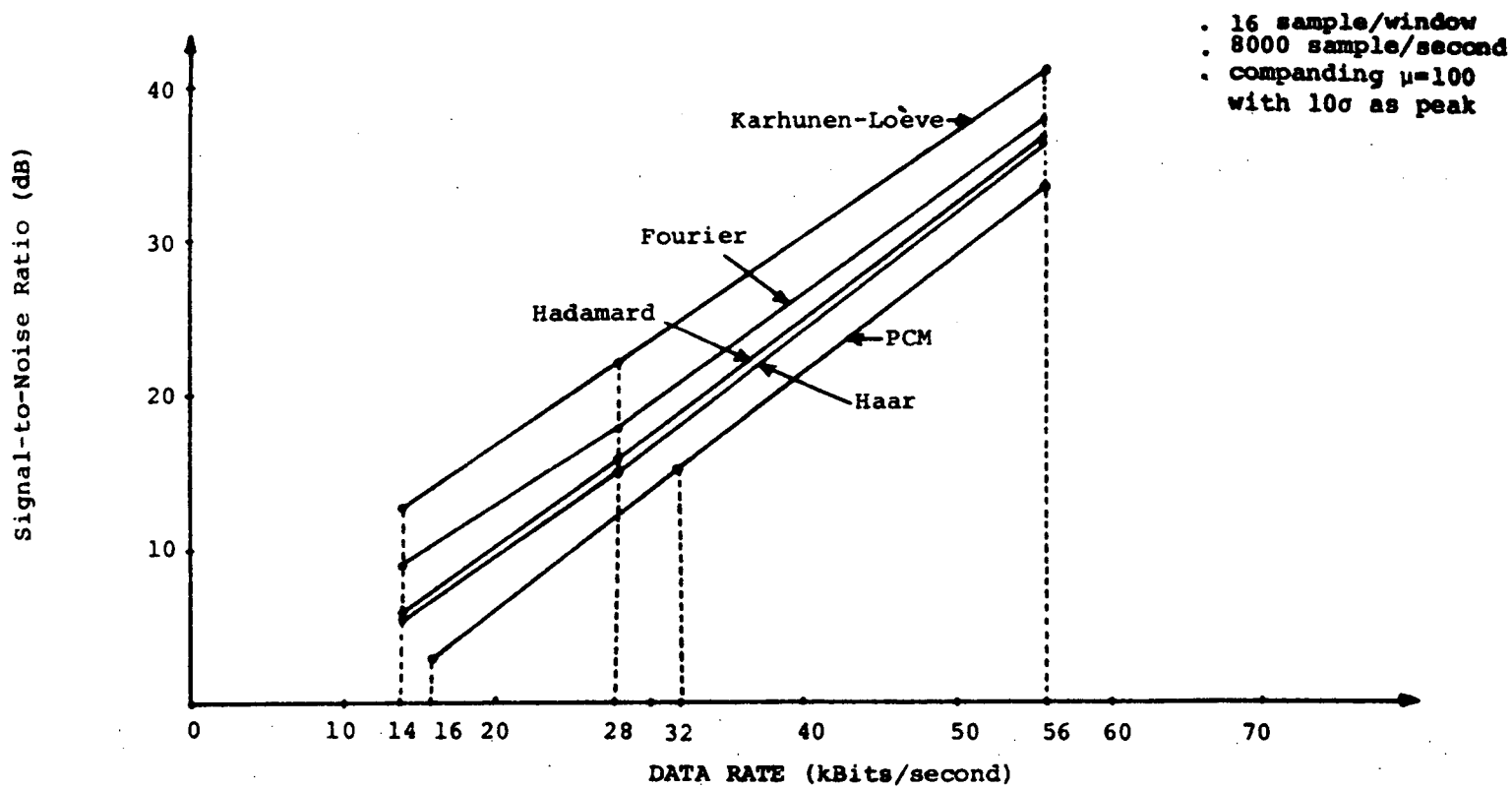


Figure 4-7. rms Error-Rate Distortion Functions for Fourier, Hadamard, Haar, and Karhunen-Loève Transforms

the improvement in signal-to-noise ratio over that provided by PCM is given.

Table 4-11. Signal-to-Quantizing Noise Ratios for Fourier, Hadamard, Haar, and Karhunen-Loève Transforms

Transform	Bit Rate (kbps)	Orthogonal Transform S/N _Q (dB)	PCM S/N _Q (dB)	$\hat{\Lambda}_Q$ Improvement in S/N _Q Over PCM (dB)
Fourier	56	37.9	33.3	4.6
	28	17.6	12.1	5.5
	14	8.9	2.9	6.0
Hadamard	56	36.1	33.3	2.8
	28	15.6	12.1	3.5
	14	5.4	2.9	2.5
Haar	56	35.9	33.3	2.6
	28	14.8	12.1	2.7
	14	5.24	2.9	2.34
Karhunen-Loève	56	41	33.3	7.7
	28	21.9	12.1	9.8
	14	12.6	2.9	9.7

Examination of the data in Table 4-11 shows that Karhunen-Loève processing clearly provides the greatest enhancement in performance, i.e., an average of 9.1 dB over PCM. Note that, at 14 kbps, it provides an S/N_Q equivalent to that of PCM at 28 kbps. It should be pointed out that this method of processing does require determination of a unique set of orthogonal functions for each talker. If a compromise set of orthogonal functions is used to accommodate several talkers, some degradation will be exacted; however, even with this degradation, the Karhunen-Loève transform may still outperform all of the others.

The second ranking transformation is the Fourier transformation, providing an average enhancement of 5.4 dB over PCM. This transformation may possess some processing advantage over the Karhunen-Loève transformation, since it is not necessary to determine special orthogonal function sets; thus, a permanently stored orthogonal function set is used. Also, the fast Fourier transform (FFT) method can be applied. No comparable process exists for the Karhunen-Loève transform; however, this should not constitute a serious limitation for the Karhunen-Loève method, since the FFT advantage is not great for the small window sizes used.

The third ranking technique is the Hadamard transformation. Its S/N_Q shows an average improvement of 2.93 dB over that of PCM. Hadamard processing is very simple, since the orthogonal functions are only two-valued, e.g., ± 1 . It is probably the most economical method for implementing an orthogonal transform method, but of course, as is seen from the results obtained here, it is not the most effective.

The fourth ranking technique is the Haar transformation. Actually, its performance is not very different from that of the Hadamard transformation, since it provides an improvement of 2.56 dB over PCM. In terms of implementation, it is also simple and its requirements would probably be comparable to those of the Hadamard transformation.

Another aspect of the performance of the various transformations was shown by the rms error rate-distortion functions in Figure 4-6. These curves, which were plotted by using the data from which Table 4-12 was derived, can be used to graphically determine the bit-rate reductions achieved by each of the processing methods. For 56, 28, and 14 kbps, the bit rate is reduced by an average of 13.5 kbps by Karhunen-Loève, 7.5 kbps by

Fourier, 5.0 kbps by Hadamard, and 4.0 kbps by Haar at a signal-to-quantizing noise ratio of dB. The bit-rate reductions appear to increase at lower values of signal-to-quantizing noise ratio.

Table 4-12. PB Word Scores (150 words, two talkers)

Bit Rate (kbps)	Word Score (percent)		
	Karhunen-Loève	Fourier	Hadamard
56	98	99	99
28	98	95	96
14	95	89	85

Monosyllabic PB words were also used to evaluate the performance of the various orthogonal processing methods. One-hundred fifty words, spoken by two talkers and evaluated by four listeners, were used. These tests show that, at 56 and 28 kbps, differences among the Karhunen-Loève, Fourier, and Hadamard processing methods cannot be observed. However, at 14 kbps, the Karhunen-Loève transform produces significantly better results than the Hadamard. The PB word test results were given in Table 4-9.

5. ORTHOGONAL TRANSFORMATION TECHNIQUES FOR TWO DIMENSIONS

5.1 INTRODUCTION

The use of conventional pulse-code modulation techniques to transmit television signals requires a very wideband transmission medium. Therefore, new bandwidth compression (or, equivalently, bit-rate reduction) techniques have been explored for digital transmission of television signals.

Reduction of the bit rate required for the transmission of a television channel is possible if the coefficients obtained from a suitable orthogonal transform of the television signal are properly coded. The advantage of using an orthogonal transformation depends on the suitability of the transformation for the representation of the signal and the simplicity of the implementation. The overall performance can then be judged by using conventional PCM as a reference.

With the development of high-speed digital computers, various transformations have recently been explored by computer simulation for possible bit-rate reduction. The well-known Fourier transform, which can be computed very rapidly, is typical of the orthogonal transforms which are being studied. Although the FFT algorithm is very efficient, its implementation is not simple. However, the Fourier transform of an image can be practically determined by using a coherent optical system.

Another transformation which can be computed very rapidly is the discrete Walsh or Hadamard transformation which does not require any multiplication operations. Alternatively, the television signal may be considered as a random process and the eigenvectors derived from the covariance matrix of the process may be used to represent the signal. This transform is known as the Karhunen-Loève transform.

These three transformations have been studied and used to process a moonscape slide at various bit rates. To produce an array of samples for image processing, the moonscape slide was scanned by a flying spot scanner. A general-purpose computer was used to perform spatial domain and transform domain processing on this array of data. The processed image was then reconstructed by the flying spot scanner.

PCM processing in the spatial domain is performed on a line-by-line basis. Orthogonal processing has been accomplished on small $N \times N$ blocks of the image frame.

The application of linear transformation and block quantization to image bandwidth reduction problems has been studied by several researchers. An extensive bibliography is included as Section 7.6.

There are two sources of error in the transmission of an analog image in digital form over a physical transmission medium: quantization error and transmission error (channel error). The digital system must be designed for the best overall performance according to a suitable performance criterion. However, because of the complexity of the overall system, the usual approach to source and channel encoding has been to consider these two errors as separate problems. This is somewhat justified, since quantization errors can be reduced by using nonlinear quantization and, alternatively, channel errors can be reduced by channel encoding. Therefore, in the study of orthogonal transforms for possible image transmission rate reduction, channel effects will be ignored.

A good understanding of the statistical properties of the orthogonal transform domain is required to develop efficient quantization and coding methods.

5.2 PCM IMAGE PROCESSING

In PCM transmission of an image,^{33,34} the continuous image is first sampled in the spatial domain to produce an $L \times L$ array of discrete samples which are quantized in brightness by using $N_b = 2^M$ levels. Then, B , the total number of bits per frame to be transmitted, is given by

$$B_t = L^2 M$$

Subjective tests are made to determine the tradeoff between L and M which will result in the "best" picture.³⁵ The result necessarily depends on the type of material and the definition of the "best" picture. For commercial applications, however, the final decision is based on subjective tests. The values for the U.S. commercial television systems are $L = 525$ and $M = 8$ bits per sample.

The image degradation which results when the quantization process uses a smaller number of bits per sample is commonly known as the contouring effect, which is the formation of steps or level changes instead of gradually changing brightness. This effect becomes perceptible when M is reduced to six or seven bits per sample. For PCM picture transmission, some of the well-known techniques for removing the effect of the quantization noise are addition of pre-emphasis and de-emphasis networks to conventional PCM,³⁶ use of pseudo-random noise,³⁷ and nonuniform quantization.^{38,39} Also, techniques have been proposed for shaping the quantizing noise spectrum by using feedback around the quantizer;^{40,41} these techniques have found application in low-resolution picture transmission systems.

5.3 QUANTIZING IN THE TRANSFORM DOMAIN

In the transform domain, the effect of quantizing is entirely different from that encountered in conventional PCM. The intensity of any one reconstructed image element in an $N \times N$ array of picture elements is the weighted average of the contributions of N^2 quantized coefficients. Thus, any one picture element does not reflect the quantization steps used to specify the coefficients. Consequently, no contouring is observed even at rates as low as two bits per picture element. As previously mentioned, Roberts' pseudo-random noise technique,³⁷ which involves the addition of noise to the signal before and after quantizing to smear out contours, has been used to reduce the subjective perception of contours. The averaging process provided by the transform domain processing is considered to be a superior means of removing such contouring, since it deliberately exploits the redundancy properties of the picture.

Transform coding makes bit-rate reduction possible because of the greater concentration of energy in the transform domain. This can be seen most clearly in terms of the Fourier transformation. When the Fourier transform of a frame of picture samples is taken and the magnitudes of the Fourier coefficients are displayed on a CRT tube, it is seen that the intensity is very high near the origin and along the vertical and horizontal axes. The value at the origin of the Fourier domain gives the average brightness of the picture, and for most pictures, the intensity distribution in the Fourier domain approximates the shape of a hyperbola. A similar situation occurs in the Hadamard transform domain.

For high-quality transformed image transmission, every coefficient must be coded in proportion to its importance to the reconstruction of the image. Thus, the bit distribution must be

in proportion to the variances of the coefficients. Since obtaining statistics of the transform coefficients is a monumental task, another approach has been used. In this approach, the picture is subdivided into smaller blocks and the variances are obtained by ensemble averaging over all of the blocks of a single frame.

This is also reasonable in terms of implementation, since transform processing a full frame in a single transformation would be considerably more involved and would probably not yield significantly improved results.

5.4 TWO-DIMENSIONAL DISCRETE FOURIER TRANSFORM (2D-DFT)

Let $f(x,y)$ denote a two-dimensional square array of values obtained by sampling the brightness of an image at $N \times N$ points. (A practical value of N in terms of processing is 16.) Let $F(u,v)$ be the two-dimensional Fourier transform of $f(x,y)$. Mathematically, such a two-dimensional discrete Fourier transform is defined as*

$$F(u,v) = \frac{1}{N_1 N_2} \sum_{x=0}^{N_1-1} \sum_{y=0}^{N_2-1} f(x,y) \exp \left[-2\pi i \left(\frac{ux}{N_1} + \frac{vy}{N_2} \right) \right]$$

$$u = 0, 1, \dots, N_1 - 1$$

$$v = 0, 1, \dots, N_2 - 1$$

$$f(x,y) = \sum_{u=0}^{N_1-1} \sum_{v=0}^{N_2-1} F(u,v) \exp \left[+2\pi i \left(\frac{ux}{N_1} + \frac{vy}{N_2} \right) \right]$$

$$x = 0, 1, \dots, N_1 - 1$$

$$y = 0, 1, \dots, N_2 - 1$$

*More generally, a two-dimensional Fourier transform pair is defined on a rectangular spatial lattice of N_1 samples in the x direction and N_2 samples in the y direction:

$$F(u,v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \exp\left[-\frac{2\pi i}{N}(ux + vy)\right]$$

$$u,v = 0, 1, \dots, N-1 \quad (5-1)$$

$$f(x,y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v) \exp\left[+\frac{2\pi i}{N}(ux + vy)\right]$$

$$x,y = 0, 1, \dots, N-1 \quad (5-2)$$

where $f(x,y)$ is the image intensity function in the spatial domain, $F(u,v)$ is the Fourier transform, and u and v are called the spatial frequencies.

The two-dimensional Fourier transform $F(u,v)$ of a spatial signal function $f(x,y)$ can be computed as two sequential one-dimensional transforms, since the Fourier kernel,

$$\exp\left[\pm\frac{2\pi i}{N}(ux + vy)\right]$$

is separable and symmetric. Thus,

$$F(u,v) = \frac{1}{N} \sum_{y=0}^{N-1} \exp\left(-\frac{2\pi i}{N} vy\right) \underbrace{\left[\frac{1}{N} \sum_{x=0}^{N-1} f(x,y) \exp\left(-\frac{2\pi i}{N} ux\right) \right]}_{\Gamma(u,y)}$$

For image processing applications, $f(x,y)$ is a positive real function representing the brightness of the spatial sample. The horizontal transform $\Gamma(u,y)$ is a complex function. However, it has a conjugate symmetry property:

$$\Gamma^*(u, y) = \Gamma(N-u, y) \quad , \quad u = 1, 2, \dots, \frac{N}{2} - 1 \quad (5-3)$$

The two-dimensional Fourier transform of a real-valued function has the following conjugate symmetry property:

$$F^*(u, v) = F(N-u, N-v) \quad , \quad u, v = 1, 2, \dots, \frac{N}{2} - 1 \quad (5-4)$$

The zero spatial frequency term is a real number and represents the two-dimensional average of the image brightness:

$$F(0, 0) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \quad (5-5)$$

For $u = v = N/2$, the two-dimensional Fourier transform is real:

$$F\left(\frac{N}{2}, \frac{N}{2}\right) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (-1)^{x+y} f(x, y) \quad (5-6)$$

This term is analogous to the folding frequency term encountered in the one-dimensional FFT analysis of a bandlimited time function. The $F(0, N/2)$ and $F(N/2, 0)$ components are real but unequal for a spatial function with no particular symmetry. That is,

$$F\left(0, \frac{N}{2}\right) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (-1)^y f(x, y)$$

$$F\left(\frac{N}{2}, 0\right) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (-1)^x f(x, y)$$
(5-7)

The complex conjugate symmetry properties of the two-dimensional discrete Fourier transform for real $f(x, y)$ are shown in Figure 5-1, where spatial frequency components which are complex conjugates are joined by straight arrows. Examination of Figure 5-1 shows that, if the magnitudes (or logarithms of the magnitudes) of the spectral components are displayed with varying intensities, the two-dimensional spectral plot thus obtained will have an odd symmetry with respect to $u = N/2$ and $v = N/2$ lines and $v = u$ and $v = -u + N - 1$ diagonals, with the exception of points on the u and v axes. The spectral points on these axes also have a certain conjugate symmetry property:

$$F(0, v) = F^*(0, N-v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \exp\left[-\frac{2\pi i}{N} vy\right]$$

$$F(u, 0) = F^*(N-u, 0) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \exp\left[-\frac{2\pi i}{N} ux\right]$$
(5-8)

The $u = N/2$ and $v = N/2$ axes have the following conjugate symmetry property:

$$F\left(\frac{N}{2}, v\right) = F^*\left(\frac{N}{2}, N-v\right)$$

$$F\left(u, \frac{N}{2}\right) = F^*\left(N-u, \frac{N}{2}\right)$$
(5-9)

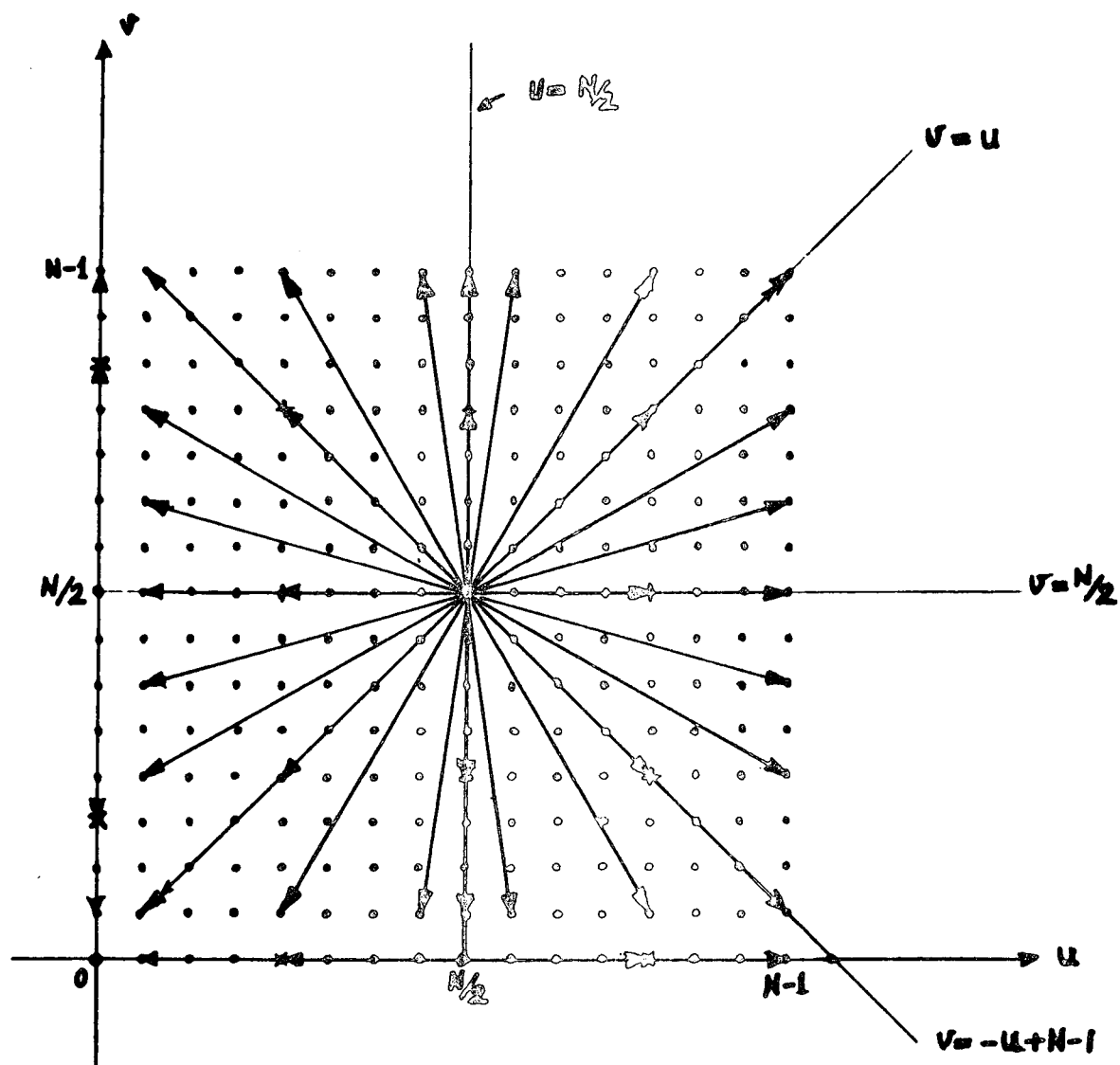


Figure 5-1. Complex Conjugate Symmetry Properties of Two-Dimensional Discrete Fourier Transform

As a result of the conjugate symmetry properties, while the transform contains $2N^2$ components (the real and imaginary or magnitude and phase components of each spatial frequency), only N^2 data components are required to describe the Fourier transform of an image. Pratt and Andrews have produced CRT displays of the Fourier transforms of some images. They have shifted the origin to the center of the transform by multiplying the image function by $(-1)^{x+y}$ before the transmission:

$$G(u,v) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} (-1)^{x+y} f(x,y) \exp \left[-\frac{2\pi i}{N} (ux + vy) \right]$$

Then it can easily be seen that

$$G(u,v) = F\left(u+\frac{N}{2}, v+\frac{N}{2}\right) \quad (5-10)$$

This shifted form is useful for display purposes only. The reconstruction of the images is accomplished by using $F(u,v)$.

A two-dimensional discrete Fourier transform has been computed over $N \times N$ ($N = 16$) blocks of the moonscape slide. First- and second-order statistics of the Fourier coefficients are computed by ensemble averaging over the 1024 blocks. Figure 5-2 plots the isovariance contours of the coefficient variances. The values of the variances are in dB because of large dynamic ranges. Each of the frequency variables u and v takes on $N = 16$ discrete values. It should be noted that the value at the origin represents the average brightness of the individual $N \times N$ block. Obviously, this value should be quantized very accurately; otherwise, a checker-board effect can occur over the full frame.

To minimize the mean-square error caused by quantization, the bit distribution in the transform domain is made in accordance

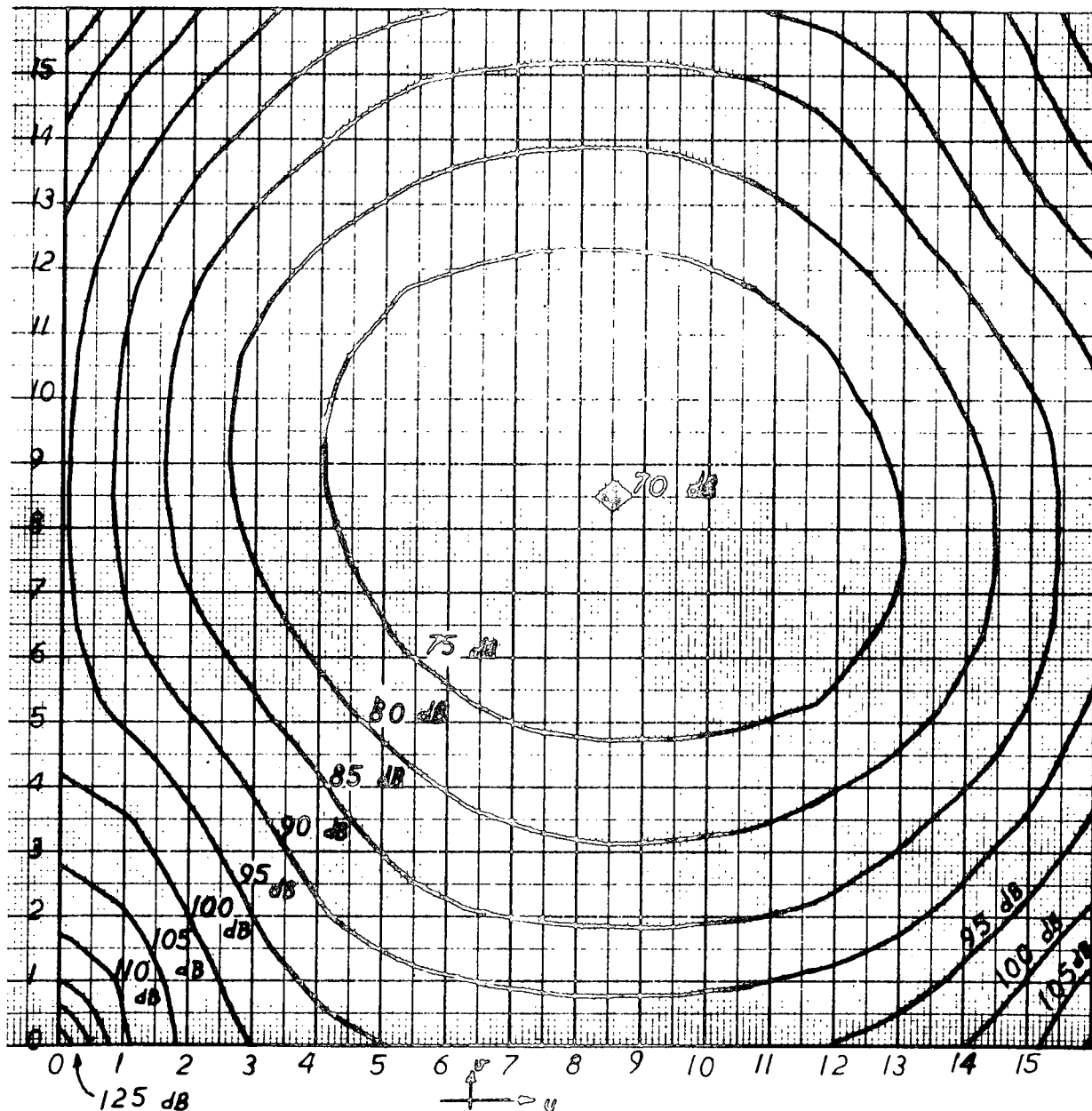


Figure 5-2. Two-Dimensional Isovariance Contours in the Fourier Domain

with the variances of the coefficients, with the constraint that the total number of bits should not exceed a given maximum. This maximum is determined by considering PCM transmission of the same image.

The total quantizing noise caused by the quantized transmission of the $N \times N$ block of image is

$$\overline{\epsilon^2} = \frac{1}{N^2} \sum_{i=0}^{N^2-1} K_i \frac{\sigma_i^2}{2^{2n_i}} \quad (5-11)$$

where K_i is a constant depending on the quantizer structure used for quantizing the i th coefficient, σ_i^2 is the variance of the i th coefficient, and n_i is the number of bits assigned to the i th coefficient.

Bit distribution in the transform domain is made so that the total quantizing noise given by equation (5-11) is minimized. The total number of bits,

$$\sum_{i=0}^{N^2-1} n_i$$

is determined by considering PCM transmission of the $N \times N$ block of image.

In PCM transmission of an image, the spatial domain samples are quantized by using M bits per sample. The total number of bits required to transmit an $N \times N$ block of frame is then

$$B = N^2 M$$

Typical values chosen for M are seven, four, and two bits per sample; thus, the total number of bits used is $B = 1792, 1024$, and

512, respectively. In order to evaluate the signal-to-quantization noise performance of discrete Fourier transform processing of the same image, the coefficients are assigned bits according to their variances and the constraint that the average number of bits,

$$\overline{n_i} = \frac{1}{N^2} \sum_{i=0}^{N^2-1}$$

must be equal to M .

Three different values are assigned to $\overline{n_i}$: 7, 4, and 2. The bit distributions for these three values are given in Tables 5-1, 5-2, and 5-3. In these tables, it should be noted that there are two numbers for each spatial frequency term. The number in the upper half is the number of bits assigned to the real part of the coefficient; the number in the lower half is the number of bits assigned to the imaginary part of the coefficient. The lower halves of four blocks are blocked, since the coefficients corresponding to these spatial frequencies are real. The zero spatial frequency term, which represents the average brightness of the image, has the highest variance; therefore, the maximum number of bits assigned appears in the upper left corner of each table. For the three cases considered, the maximum numbers of bits are 15, 12, and 10, respectively. The moonscape data has been processed at three different bit rates using the bit distribution given in Tables 5-1, 5-2, and 5-3. The resulting images are shown at the end of this report.

A uniform quantization rule has been used for quantizing the coefficients. The quantizer spread, Q_S (defined as the distance between the minimum and maximum levels of the quantizer) is given by

$$Q_S = 2\alpha\sigma$$

Table 5-1. Optimum Bit Distribution in Two-Dimensional
Complex Fourier Domain
(average seven bits per sample)

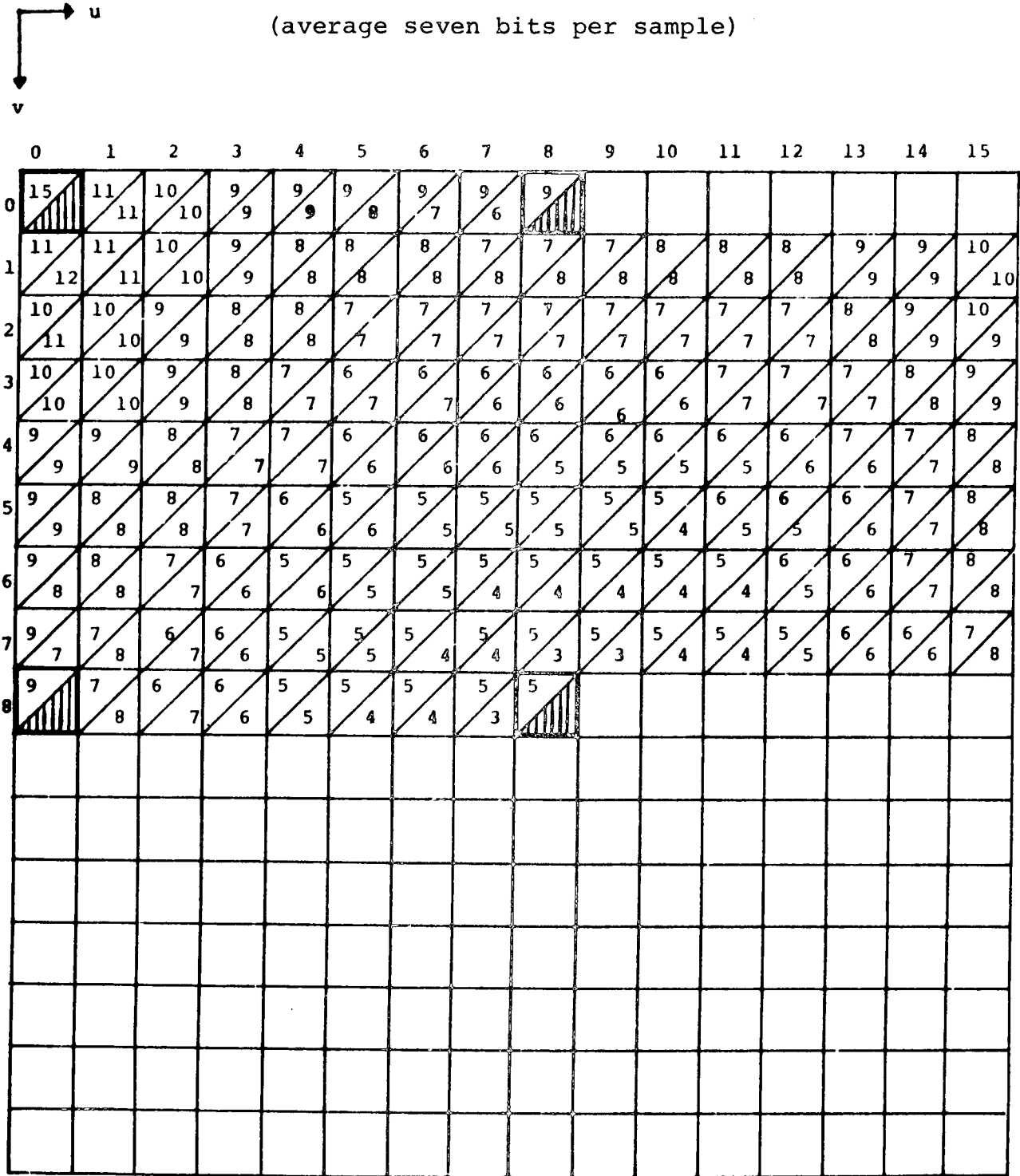
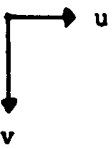


Table 5-2. Optimum Bit Distribution in Two-Dimensional
Complex Fourier Domain


(average four bits per sample)



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	12	8	7	6	6	6	6	6	6							
1	8	8	7	6	5	5	5	4	4	4	5	5	5	6	6	7
2	7	7	6	5	5	4	4	4	4	4	4	4	4	5	6	7
3	7	7	6	5	4	3	3	3	3	3	3	3	4	4	5	6
4	6	6	5	4	4	3	3	3	3	3	2	2	3	4	4	5
5	6	5	5	4	3	2	2	2	2	2	2	1	2	3	4	5
6	6	5	4	3	2	2	2	2	2	2	1	1	2	3	4	5
7	6	4	3	3	2	2	2	2	2	2	1	1	2	3	3	4
8	6	4	3	3	2	2	2	2	2	2	1	1	2	3	3	4

Table 5-3. Optimum Bit Distribution in Two-Dimensional
Complex Fourier Domain

(average two bits per sample)



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	10	6	4	4	4	4	4	4	4							
1	6	6	5	4	3	3	2	2	2	2	3	3	3	4	4	5
2	5	5	4	3	3	2	2	1	1	1	2	2	2	3	4	5
3	5	5	4	3	2	1	1	1	1	1	1	1	2	2	3	4
4	4	4	3	2	2	1	1	1	0	0	1	1	1	1	2	3
5	4	3	2	2	1	0	0	0	0	0	0	0	1	1	2	3
6	4	3	2	1	0	0	0	0	0	0	0	0	0	1	1	2
7	4	2	1	1	0	0	0	0	0	0	0	0	0	1	1	2
8	4	2	1	0	0	0	0	0	0							

where σ is the variance of the coefficient, and α is the amplitude loading factor of the quantizer. It is known that α should be dependent on the number of bits assigned to the coefficient in order to minimize the quantization noise. Hence, the quantizer spread would be made adaptable. However, as indicated by an example in Section 6, the resulting additional improvement in signal-to-quantizing noise ratio is not significant. Therefore, for a given data rate, α is chosen to be the same for each coefficient. At various stages of the processing performed in this study, various values of α have been used.

For PCM processing of the image, the quantizer is uniform; hence, the quantizer spread is equal to the full range between zero intensity (black) and maximum intensity (white).

The error caused by intensity quantization has been computed by comparing the input and output intensity values for each line of the entire frame. For the discrete Fourier transform, the error caused by quantization of the coefficients has been computed by comparing the input and output intensity values over each block of the entire frame. The results will be given in Section 6.

5.5 TWO-DIMENSIONAL DISCRETE WALSH-HADAMARD TRANSFORMATION (2D-DWT)

Let $f(x,y)$ represent a two-dimensional square array of values obtained by sampling the brightness of an image at $N \times N$ points. (The value used for N is 16.) Let $F(j,k)$ be the two-dimensional Walsh-Hadamard transform of $f(x,y)$. Mathematically, such a two-dimensional discrete Walsh transform pair is defined as

$$F(j,k) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \text{ wal}(j,x) \text{ wal}(k,y)$$

$$j,k = 0, 1, \dots, N-1 \quad (5-12)$$

$$f(x,y) = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} F(j,k) \text{ wal}(j,x) \text{ wal}(k,y)$$

$$x,y = 0, 1, \dots, N-1 \quad (5-13)$$

where $\text{wal}(j,x)$ and $\text{wal}(k,y)$ represent the one-dimensional Walsh functions in the x and y directions, and j and k are the numbers of zero crossings of the Walsh functions in the x and y directions, respectively. The coefficient corresponding to $j = k = 0$,

$$F(0,0) = \frac{1}{N^2} \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y)$$

yields the average brightness of the $N \times N$ block and is identical to the zero spatial frequency term in the two-dimensional discrete Fourier domain given by equation (5-5).

The application of the two-dimensional Hadamard transform can be visualized as the comparison (integral of the product) of the brightness pattern of each 16×16 image block with the various two-dimensional Hadamard basis patterns for $N = 16$. Hadamard basis patterns for $N = 8$ are shown in Figure 5-3. The coefficient corresponding to the Hadamard basis $\text{wal}(1,x) \text{ wal}(0,y)$ represents the comparison of the brightness of the left and right halves of the $N \times N$ block. The coefficient corresponding to the basis $\text{wal}(0,x) \text{ wal}(1,y)$ represents the comparison of the brightness of the lower and upper halves of the image block. Finally,

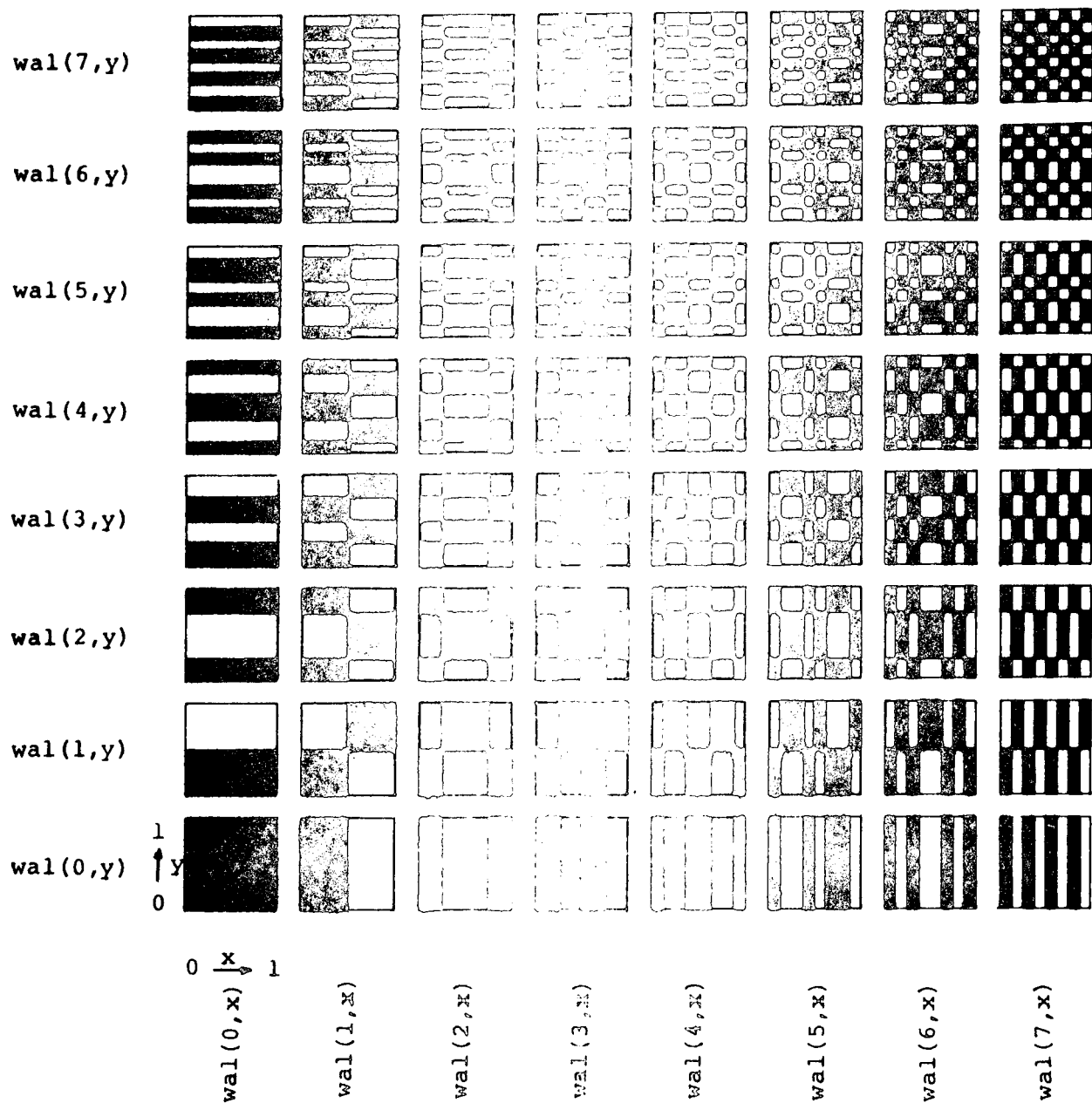


Figure 5-3. Two-Dimensional Walsh-Hadamard Basis for $N = 8$
(black areas represent $+1/N$ and white areas represent $-1/N$)

the coefficient corresponding to the $wal(N-1,x)$ $wal(N-1,y)$ basis represents the comparison of the image block with an $N \times N$ checker-board design, as shown in the upper right corner of Figure 5-3. This coefficient is identical to the spatial folding frequency term, $F(N/2,N/2)$, in two-dimensional discrete Fourier transform processing.

Since the transformation kernel,

$$wal(j,x) wal(k,y)$$

is separable, the transform can be applied as two consecutive one-dimensional transforms. Therefore, transformation coefficients of each block are computed by applying the one-dimensional Walsh transform, first in the x direction to every line, and then in the y direction to every column of the first transformation. If the Hadamard transform in natural order is used in the x direction, the coefficients should be sequency ordered before it is applied in the y direction. The resulting $N \times N$ coefficients, which represent the $N \times N$ block of image in the Walsh domain, are quantized by using a certain bit distribution, and the processed image is reconstructed as two consecutive one-dimensional transforms, first in the k direction and then in the j direction.

Bit distribution in the Walsh-Hadamard domain is determined by coefficient variances which are computed by ensemble averaging over the total number of blocks. For the three cases considered, the total numbers of bits are equal to 1792, 1024, and 512, resulting in averages of 7, 4, and 2 bits per sample, respectively. The bit distributions for these three cases are given in Tables 5-4, 5-5, and 5-6, respectively.

The zero spatial sequency (or frequency) term, $F(0,0)$, has the highest variance. Therefore, it is assigned the highest number of quantization levels in order to prevent the appearance


Table 5-4. Optimum Bit Distribution in Two-Dimensional
Walsh-Hadamard Domain

(average seven bits per sample)

Diagram showing the relationship between indices k and j :

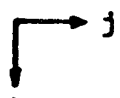
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	14	11	10	10	9	9	9	9	7	7	7	7	8	8	8	8
1	12	11	10	10	9	9	9	8	7	7	7	7	7	8	8	7
2	11	10	10	9	8	9	9	8	6	7	7	7	7	7	7	7
3	11	10	9	9	8	8	8	8	6	7	7	7	7	7	7	6
4	10	9	9	9	8	8	8	7	6	6	7	7	7	7	7	6
5	10	9	9	9	8	8	8	7	6	6	6	7	7	7	7	6
6	10	9	9	8	8	8	8	7	6	6	6	6	6	6	7	6
7	9	9	8	8	7	7	7	7	6	6	6	6	6	6	6	5
8	8	8	8	7	7	7	7	6	5	6	6	6	6	6	6	5
9	8	8	8	7	7	7	7	6	5	6	6	6	6	6	5	5
10	8	8	8	7	7	7	7	6	5	6	6	6	6	6	6	5
11	9	8	8	7	7	7	7	6	5	6	6	6	6	6	5	5
12	9	8	8	7	7	7	7	6	5	5	6	6	6	6	6	5
13	9	8	8	7	7	7	7	6	5	5	5	5	5	6	5	5
14	9	8	7	7	6	6	6	6	5	5	5	5	5	5	5	5
15	8	7	7	6	6	6	6	5	4	4	4	4	5	4	4	4

Table 5-5. Optimum Bit Distribution in Two-Dimensional
Walsh-Hadamard Domain
(average four bits per sample)



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	11	8	7	7	6	6	6	6	4	4	4	4	5	5	5	5
1	9	8	7	7	6	6	6	5	4	4	4	4	4	5	5	4
2	8	7	7	6	5	6	6	5	3	4	4	4	4	4	4	4
3	8	7	6	6	5	5	5	5	3	4	4	4	4	4	4	3
4	7	6	6	6	5	5	5	4	3	3	4	4	4	4	4	3
5	7	6	6	6	5	5	5	4	3	3	3	4	4	4	4	3
6	7	6	6	5	5	5	5	4	3	3	3	3	3	3	4	3
7	6	6	5	5	4	4	4	4	3	3	3	3	3	3	3	2
8	5	5	5	4	4	4	4	3	2	3	3	3	3	3	3	2
9	5	5	5	4	4	4	4	3	2	3	3	3	3	3	2	2
10	5	5	5	4	4	4	4	3	2	2	3	3	3	3	3	2
11	6	5	5	4	4	4	4	3	2	2	3	3	3	3	2	2
12	6	5	5	4	4	4	4	3	2	2	2	3	3	3	3	2
13	6	5	5	4	4	4	4	3	2	2	2	2	2	3	2	2
14	6	5	4	4	3	3	3	3	2	2	2	2	2	2	2	2
15	5	4	4	3	3	3	3	2	1	1	1	2	2	1	1	1

Table 5-6. Optimum Bit Distribution in Two-Dimensional
Walsh-Hadamard Domain
(average two bits per sample)



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	9	6	5	5	4	4	4	4	2	2	2	2	3	3	3	3
1	7	6	5	5	4	4	4	3	2	2	2	2	2	2	3	2
2	6	5	5	4	3	4	3	3	1	2	2	2	2	2	2	2
3	6	5	4	4	3	3	3	3	1	2	2	2	2	2	2	1
4	5	4	4	4	3	3	3	2	1	1	2	2	2	2	2	1
5	5	4	4	4	3	3	3	2	1	1	1	2	2	2	2	1
6	5	4	4	3	3	3	3	2	1	1	1	1	1	1	1	1
7	4	4	3	3	2	2	2	2	1	1	1	1	1	1	1	0
8	3	3	3	2	2	2	2	1	0	1	1	1	1	1	0	0
9	3	3	3	2	2	2	2	1	0	1	1	1	1	1	0	0
10	3	3	3	2	2	2	2	1	0	0	1	1	1	1	1	0
11	4	3	3	2	2	2	2	1	0	0	1	1	1	1	0	0
12	4	3	3	2	2	2	2	1	0	0	0	1	1	1	0	0
13	4	3	3	2	2	2	2	1	0	0	0	0	0	0	0	0
14	4	3	2	2	1	1	1	1	0	0	0	0	0	0	0	0
15	3	2	2	1	1	1	0	0	0	0	0	0	0	0	0	0

of block structure in the reconstructed image. The number of bits assigned to $F(0,0)$ appears in the upper left corner of each table. For the three cases considered, the numbers of bits assigned to $F(0,0)$ are 14, 11, and 9 bits, respectively.

A uniform quantization rule has been used for quantizing the coefficients. As in the case of two-dimensional discrete Fourier transform processing, the amplitude loading factor of the uniform quantizer is chosen to be a constant. The error caused by quantization of the coefficients has been computed by comparing the input and output intensities over each block of the entire frame. The results will be given in Section 6.

5.6 TWO-DIMENSIONAL DISCRETE KARHUNEN-LOÈVE TRANSFORM (2D-KLT)

To obtain a two-dimensional discrete Karhunen-Loève transform of an $N \times N$ block of image, the $N \times N$ array of sampled brightness values $f(x,y)$ of the block must be expanded in terms of the two-dimensional eigenfunction $\phi_{j,k}(x,y)$; i.e.,

$$f(x,y) = \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} F(j,k) \phi_{j,k}(x,y)$$
$$x,y = 0, 1, \dots, N-1 \quad (5-14)$$

where N^2 is the number of eigenfunctions that is necessary to adequately represent the random image signal, $f(x,y)$. The expansion coefficients, $F(j,k)$, are the eigenvalues of the $N^2 \times N^2$ spatial covariance matrix, $C(x_1, x_2, y_1, y_2)$ of $f(x,y)$, where x_1, x_2, y_1 , and y_2 are spatial variables. The form of the eigenvectors $\phi_{j,k}(x,y)$ depends on the statistical nature of the random signal; in general, it is difficult to describe their exact analytical form.

If all of the eigenvectors are used in the expansion, then the mean-square error between the original $f(x,y)$ and that of equation (5-14) is zero. If only $N_1 \leq N^2$ eigenvectors are used, then the mean-square error is given by

$$\overline{\epsilon^2} = C(0,0,0,0) - \frac{1}{N^2} \sum_{i=0}^{N_1-1} \sigma_i^2 \quad (5-15)$$

where $C(0,0,0,0)$ is the value of the spatial domain autocorrelation function at the origin, and σ_i^2 are the variances of the transform domain coefficients.

The orthogonal transform which results in the greatest reduction of the mean-square error in equation (5-15) is the Karhunen-Loève transform, in which the eigenvectors are arranged to correspond to the eigenvalues in descending order. The variances of the coefficients are the eigenvalues of $C(x_1, x_2, y_1, y_2)$.

When N is large, computation of the eigenvalues of the $N^2 \times N^2$ spatial covariance matrix becomes formidable. However, if the two-dimensional autocorrelation function is separable in the x and y directions, then the Karhunen-Loève transform can be computed as two successive one-dimensional linear transformations in the x and y directions. The transforms for the x and y directions are based on the horizontal and vertical correlations of the image. If the correlations in the x and y directions are the same for a particular image, then a two-dimensional KLT is obtained by using the KLT matrix based on the x -dimension covariance matrix to transform all of the rows of the image block, and by then performing the same transform in the y direction using the same matrix. The N^2 coefficients thus obtained are uncorrelated; their variances are given by the products of the eigenvalues of the covariance matrices representing the correlations in the x and y directions.

The application of KLT thus requires the computation of autocorrelation functions $R_X(x_1, x_2)$ and $R_Y(y_1, y_2)$.

Experimental evidence indicates that, for most imagery data, a reasonable model of the one-dimensional autocorrelation function is^{42, 43}

$$R(x_1, x_2) = R(0) e^{-\alpha |x_1 - x_2|} \quad (5-16)$$

which is equivalent to representing the signal as a first-order Markov process. In equation (5-16), $R(0) = \sigma_s^2$ is the variance of the signal, and α is a constant. As α changes, the mathematical model will fit images with different average detail.

In the discrete case, the sampled imagery signal can be modeled as a stationary first-order Markov sequence:

$$s_n = \rho s_{n-1} + \xi_{n-1} \quad (5-17)$$

where

ξ = white Gaussian input, $N(0, \sigma^2)$
 s = output, first-order Markov process
 ρ = correlation coefficient, $|\rho| < 1$

The autocorrelation sequence corresponding to equation (5-17) is

$$R(k - j) = R(0) \rho^{|k-j|}, \quad |k - j| = 1, 2, 3, \dots \quad (5-18)$$

where

$$R(0) = \sigma_s^2 = \frac{\sigma^2}{1 - \rho^2} \quad (5-19)$$

is the variance of the signal sequence. Comparison of equations (5-16) and (5-18) yields

$$\alpha = -\ln \rho \quad (5-20)$$

in order to fit the continuous and discrete Markov process models with the same picture detail.

For most imagery data, there is a high correlation between two consecutive samples in the spatial domain.^{44, 45} The correlations in the horizontal and vertical directions are not necessarily the same. This requires a two-dimensional model given by

$$R(x_1, x_2, y_1, y_2) = R(0, 0) \left[e^{-\alpha |x_1 - x_2| - \beta |y_1 - y_2|} \right] \quad (5-21)$$

The normalized covariance matrix of the first-order Markov process is given by

$$\frac{E\{s_k s_j\}}{E\{s_k^2\}} = \rho^{|k-j|} \quad , \quad k, j = 0, 1, 2, \dots, N-1 \quad (5-22)$$

which is a Toeplitz matrix of a very special form:

$$C_s = \begin{bmatrix} 1 & \rho & \rho^2 & . & . & . & \rho^{N-1} \\ \rho & 1 & \rho & . & . & . & \rho^{N-2} \\ \rho^2 & \rho & 1 & & & & . \\ . & & & . & & & . \\ . & & & & . & & . \\ . & & & & & 1 & \rho \\ \rho^{N-1} & \rho^{N-1} & . & . & . & \rho & 1 \end{bmatrix} \quad (5-23)$$

The Karhunen-Loève transform matrix is the transpose of the matrix which diagonalizes the covariance matrix. Thus, the rows of the KLT matrix are the eigenvectors of the process. The eigenfunctions are ordered so that corresponding eigenvalues decrease in value. That is,

$$\lambda_0 \geq \lambda_1 \geq \lambda_2 \dots \geq \lambda_{N-1} \quad (5-23)$$

The normalized autocorrelation sequence $\rho^{|k-j|}$ is shown in Figure 5-4 for various values of ρ . These values have been used to form the covariance matrix and to compute eigenvalues and eigenvectors. The eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_N$ are shown in Tables 5-7 and 5-8 for various N . In general, as N increases, the value of λ increases monotonically. However, if ρ is small (e.g., $\rho = 0.5$), the value of λ saturates as N increases. It should also be noted that

$$\sum_{i=0}^{N-1} \lambda_i = N \quad (5-24)$$

The sinusoidal but nonharmonic nature of the eigenfunctions of a first-order Markov process was first reported by Davenport and Root.²² Their results can be used to compute the eigenfunctions of the first-order continuous Markov process for various amounts of correlation. This involves finding the roots of two transcendental equations.⁴⁶ In the discrete case, the eigenfunctions of the first-order Markov process are samples of cosine and sine functions.⁴⁷ Nevertheless, since these sinusoidal functions are not harmonically related, the Karhunen-Loève and discrete Fourier transforms are not identical even for large N .*

* $N = 2$ is the degenerate case in which KLT is identical to discrete Fourier, Walsh, and Haar transforms, no matter what the statistical nature of the process is.

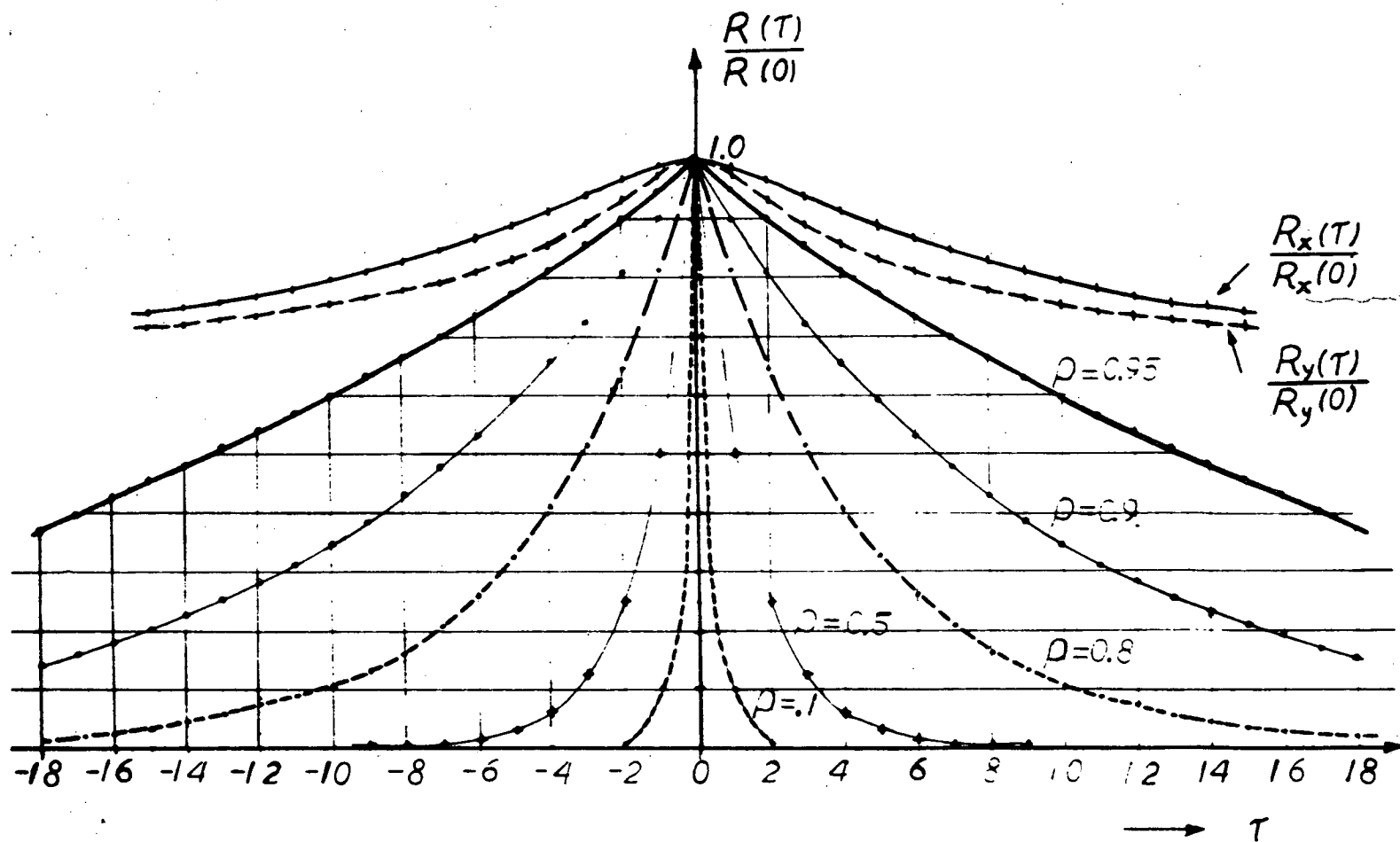


Figure 5-4. Normalized Autocorrelation Function of the First-Order Markov Process and Vertical and Horizontal Autocorrelation Functions of the Moonscape

Table 5-7. Eigenvalues of the First-Order Markov Process ($\rho = 0.9$)

λ	Size of the Covariance Matrix									
	N = 2	N = 3	N = 4	N = 5	N = 6	N = 7	N = 8	N = 10	N = 12	N = 16
λ_0	1.9	2.74	3.52	4.26	4.95	5.59	6.20	7.30	8.28	9.92
λ_1	0.1	0.19	0.31	0.45	0.62	0.80	1.00	1.44	1.92	2.94
λ_2	.	0.06	0.10	0.14	0.19	0.26	0.32	0.49	0.68	1.12
λ_3	.	.	0.06	0.07	0.10	0.13	0.16	0.24	0.33	0.56
λ_4	.	.	.	0.05	0.06	0.08	0.10	0.14	0.20	0.34
λ_5	0.05	0.06	0.07	0.10	0.13	0.22
λ_6	0.05	0.06	0.07	0.10	0.16
λ_7	0.05	0.06	0.08	0.12
λ_8	0.05	0.06	0.10
λ_9	0.05	0.06	0.08
λ_{10}	0.05	0.07
λ_{11}	0.05	0.06
λ_{12}	0.06
λ_{13}	0.05
λ_{14}	0.05
λ_{15}	0.05

Table 5-8. Eigenvalues of the First-Order Markov Process ($\rho = 0.5$)

λ	Size of the Covariance Matrix									
	N = 2	N = 3	N = 4	N = 5	N = 6	N = 7	N = 8	N = 10	N = 12	N = 16
λ_0	1.5	1.84	2.08	2.26	2.39	2.49	2.57	2.68	2.75	2.84
λ_1	0.5	0.75	1.00	1.23	1.43	1.61	1.76	2.01	2.20	2.45
λ_2	.	0.40	0.53	0.69	0.84	1.00	1.14	1.41	1.64	1.99
λ_3	.	.	0.37	0.45	0.55	0.66	0.77	1.00	1.21	1.58
λ_4	.	.	.	0.36	0.41	0.48	0.56	0.73	0.91	1.29
λ_5	0.35	0.39	0.44	0.57	0.71	1.00
λ_6	0.34	0.37	0.46	0.57	0.81
λ_7	0.34	0.40	0.48	0.68
λ_8	0.36	0.42	0.58
λ_9	0.34	0.38	0.50
λ_{10}	0.35	0.45
λ_{11}	0.33	0.41
λ_{12}	0.38
λ_{13}	0.35
λ_{14}	0.34
λ_{15}	0.33

Figure 5-5 shows the eigenvectors of the Markov process for $N = 16$ and $N = 8$. All eigenvectors have been normalized to have unity magnitude. As ρ increases, the first eigenvector, which corresponds to the largest eigenvalue, approaches the average value of N samples. Higher index eigenfunctions approximate a set of periodic sine and cosine functions. Their nonharmonic nature is evident when they are compared with the harmonic sinusoidal functions shown in the last column of Figure 5-5. For example, ϕ_1 has one-half cycle in the unit interval, ϕ_2 has one cycle, and ϕ_3 has one and one-half cycles.

The horizontal and vertical autocorrelation functions, $R_X(\tau)/R_X(0)$ and $R_Y(\tau)/R_Y(0)$, respectively, have been computed for the moonscape slide as an ensemble average of autocorrelation functions over the lines and columns of the image frame. These functions have been shown in Figure 5-4.

The covariance matrices $C_X(\tau)$ and $C_Y(\tau)$ are formed by using a method which is similar to that used for the one-dimensional case, and their eigenvalues and eigenvectors are computed. The eigenvalues are shown in Table 5-9 and the eigenvectors of $C_X(\tau)$ are shown in Figure 5-5. The eigenvectors $C_Y(\tau)$ are not shown, since they have almost the same form. Bit distribution is made in accordance with the eigenvalues.

As in the Walsh-Hadamard and Fourier cases, the moonscape slide has been processed at three different bit rates using the one-dimensional Karhunen-Loève transform matrix based on the horizontal correlations for both horizontal and vertical transforms. Tables 5-10, 5-11, and 5-12 show the bit distributions in the transform domain. The resulting processed images appear at the end of this report.

As a further experiment, the moonscape slide has been processed at a rate of two bits per picture element by using the one-dimensional KLT matrix based on horizontal correlations for

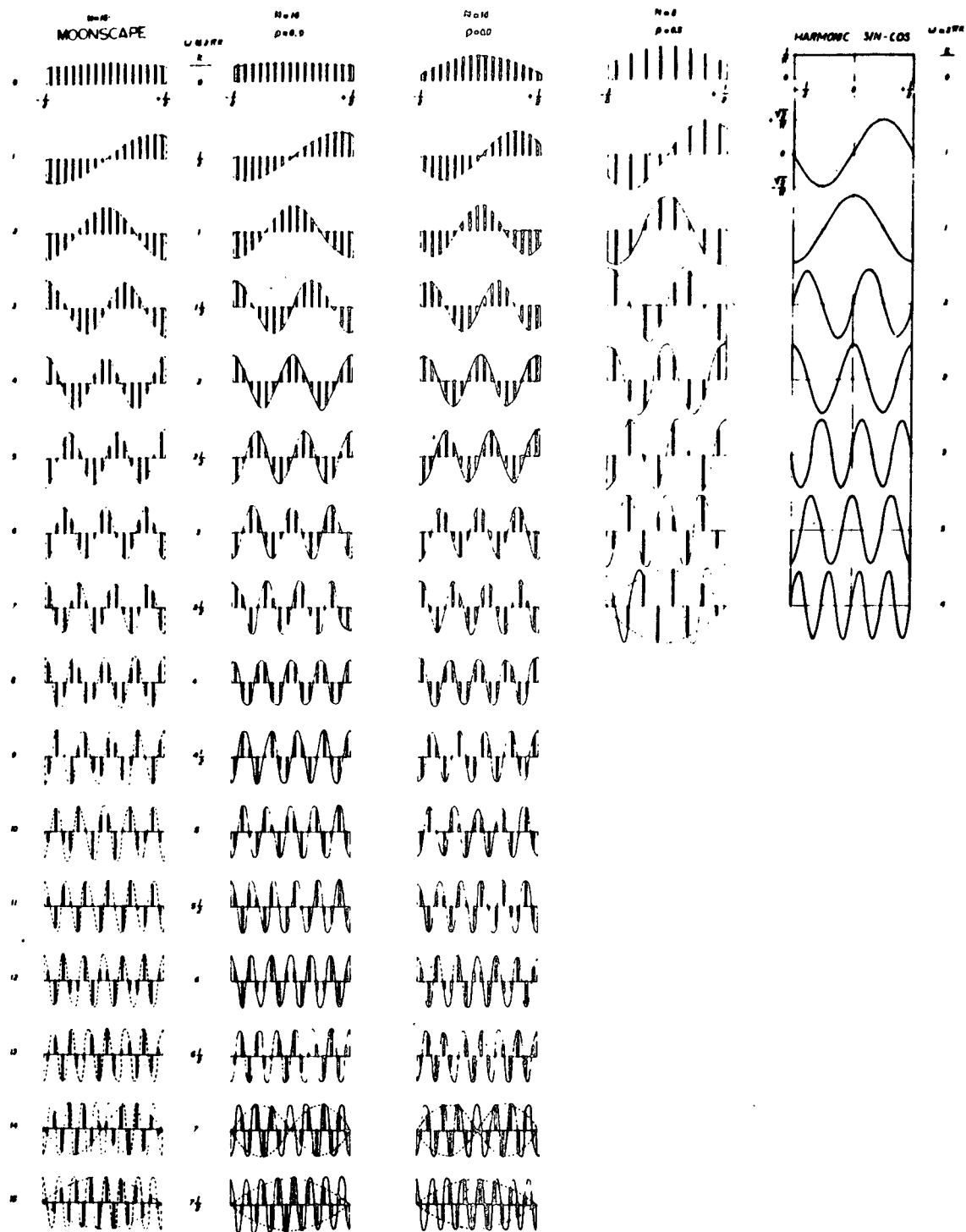



Figure 5-5. Eigenvectors of the Moonscape Covariance Matrix Based on the Horizontal Correlations and Eigenvectors of the First-Order Markov Process as Compared with Harmonic Sinusoidal Functions

horizontal transform and the transform matrix based on vertical correlations for vertical transform. The resulting bit distribution is the same as that of Table 5-12. The resulting signal-to-quantizing noise ratio is the same as it was when the same transform matrix was used for both directions.

Table 5-9. Eigenvalues of Spatial Autocovariance Matrices Representing the Correlations in the x and y Dimensions for Moonscape Data


j or k	Eigenvalues of $C_x(\tau)$, λ_j	Eigenvalues of $C_y(\tau)$, λ_k
0	14.6334	14.1362
1	0.8320	0.9052
2	0.2993	0.4393
3	0.1156	0.2150
4	0.0567	0.1249
5	0.0287	0.0754
6	0.0148	0.0464
7	0.0077	0.0266
8	0.0042	0.0150
9	0.0022	0.0080
10	0.0015	0.0043
11	0.0012	0.0023
12	0.0009	0.0011
13	0.0007	0.0004
14	0.0006	0.0002
15	0.0005	0.0001

Table 5-10. Optimum Bit Distribution in Two-Dimensional
Karhunen-Loève Transform Domain
(average seven bits per sample)



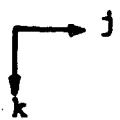
k	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	15	13	12	11	10	10	9	9	8	8	7	7	6	6	5	5
1	13	12	11	11	10	10	9	9	8	7	7	6	6	6	5	5
2	12	12	11	11	10	10	9	8	8	7	7	6	6	5	5	5
3	12	11	11	10	10	9	9	8	8	7	7	6	6	6	5	5
4	11	11	11	10	10	9	9	8	8	7	7	6	6	5	5	5
5	11	11	10	10	9	9	8	8	8	7	7	6	6	5	5	5
6	10	10	10	10	9	9	8	8	7	7	6	6	5	5	5	5
7	10	10	10	9	9	8	8	8	7	7	6	6	5	5	5	4
8	10	9	9	9	9	8	8	7	7	7	6	6	5	5	4	4
9	9	9	9	9	8	8	7	7	7	6	6	5	5	5	4	4
10	9	8	8	8	8	7	7	7	6	6	6	5	5	5	4	4
11	8	8	8	8	7	7	7	6	6	6	5	5	5	4	4	4
12	7	7	7	7	7	7	6	6	6	5	5	5	4	4	4	4
13	7	7	7	7	7	6	6	6	5	5	5	4	4	4	4	4
14	7	7	6	6	6	6	6	5	5	5	5	4	4	4	4	4
15	6	6	6	6	6	6	5	5	5	4	4	4	4	4	4	4

Table 5-11. Optimum Bit Distribution in Two-Dimensional
Karhunen-Loève Transform Domain
(average four bits per sample)



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	12	10	9	8	7	7	7	6	6	5	5	4	4	3	2	2
1	10	9	8	8	7	7	6	6	5	4	4	3	3	3	2	2
2	9	9	8	8	7	7	6	5	5	4	4	3	3	2	2	2
3	9	8	8	7	7	6	6	5	5	4	4	3	3	3	2	2
4	8	8	8	7	7	6	6	5	5	4	4	3	3	3	2	2
5	8	8	7	7	6	6	5	5	5	4	4	3	3	2	2	2
6	7	7	7	7	6	6	5	5	4	4	3	3	2	2	2	2
7	7	7	7	6	6	5	5	5	4	4	3	3	2	2	2	1
8	7	6	6	6	6	5	5	4	4	4	3	3	2	2	2	1
9	6	6	6	6	5	5	4	4	4	3	3	2	2	2	1	1
10	6	5	5	5	5	4	4	4	3	3	3	2	2	2	1	1
11	5	5	5	5	4	4	4	3	3	3	2	2	2	1	1	1
12	4	4	4	4	4	4	3	3	3	2	2	2	1	1	1	1
13	4	4	4	4	4	3	3	3	2	2	2	1	1	1	1	1
14	4	4	3	3	3	3	3	2	2	2	2	1	1	1	1	1
15	3	3	3	3	3	3	2	2	2	1	1	1	1	1	1	1

Table 5-12. Optimum Bit Distribution in Two-Dimensional
Karhunen-Loève Transform Domain
(average two bits per sample)



	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	10	7	7	6	5	5	4	3	3	2	2	1	1	1	0	0
1	8	7	6	6	5	4	4	3	3	2	2	1	1	1	0	0
2	7	7	6	5	5	4	4	3	3	2	2	1	1	0	0	0
3	7	6	6	5	5	4	4	3	3	2	2	1	1	0	0	0
4	6	6	5	5	4	4	4	3	3	2	2	1	1	0	0	0
5	6	6	5	5	4	4	3	3	2	2	1	1	1	0	0	0
6	5	5	5	4	4	4	3	3	2	2	1	1	0	0	0	0
7	5	5	4	4	4	3	3	2	2	1	1	1	0	0	0	0
8	4	4	4	4	3	3	3	2	2	1	1	0	0	0	0	0
9	4	4	4	3	3	3	2	2	1	1	1	0	0	0	0	0
10	3	3	3	3	3	2	2	1	1	1	0	0	0	0	0	0
11	3	3	3	3	2	2	2	1	1	1	0	0	0	0	0	0
12	2	2	2	2	2	2	1	1	0	0	0	0	0	0	0	0
13	2	2	2	2	1	1	1	1	0	0	0	0	0	0	0	0
14	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0
15	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0

6. ORTHOGONAL TRANSFORM PROCESSING OF IMAGES

6.1 GENERAL

A moonscape slide has been scanned by a flying spot scanner to produce an array of $L \times L$ ($L = 512$) uniformly spaced samples of image brightness. Next, a general-purpose computer has been used to perform spatial domain or transform domain processing on this array of data. Then, the processed image has been reconstructed by the flying spot scanner.

Orthogonal transformation processing has been accomplished on small $N \times N$ ($N = 16$) blocks. Thus there are $p = (512 \times 512) / (16 \times 16) = 1024$ such blocks. The orthogonal transform of each block has been computed to obtain the coefficients which represent the image in the transform domain.

6.2 FUNDAMENTALS OF THE PROCESSING METHOD

Thresholding magnitudes of the coefficients in the transform domain makes bit-rate reduction possible. That is, the energy of each coefficient is checked against a threshold; the coefficient is transmitted if its energy is above the threshold and not transmitted if its energy is below the threshold. However, this method appears to be overly crude, as shown for the discrete Fourier transform in the supplementary report no. 1 on image processing.

For high-quality image transmission, every coefficient must be considered in terms of its importance in image reconstruction. Since the amount of information contained in each coefficient is proportional to its energy, bit distribution and hence

allotment of the quantization levels should be made in accordance with the variances of the coefficients. This procedure also minimizes the rms error at any given bit rate, as shown in Section 4.2.

Accordingly, the coefficients are quantized by using bit distributions based on their variances. At the receiving side, the spatial domain samples are reconstructed from the quantized coefficients by using an inverse orthogonal transformation.

In PCM transmission of an image, the spatial domain samples are quantized by using M bits per sample. The total number of bits required to transmit an $N \times N$ block of frame is then

$$B = N^2 M$$

Typical values chosen for M are seven, four, and two bits per sample; thus the total numbers of bits used for $N = 16$ are $B = 1792$, 1024 , and 512 , respectively. In order to evaluate the signal-to-quantization noise performance of the orthogonal transform processing on the same image, the coefficients are assigned bits according to their variances and the constraint that the total numbers of bits are 1792 , 1024 , and 512 , resulting in averages of seven, four, and two bits per picture element, respectively. In Section 5, the bit distributions for these three cases were given in tables for each transform method.

6.2.1 Coefficient Quantization

In order to obtain the maximum signal-to-quantizing noise ratio, the amplitude loading factor of the uniform quantizer in the transform domain should be dependent on the number of bits assigned to the coefficients. If σ_i^2 is the variance of the i th coefficient, the quantizer spread (defined as the distance between

minimum and maximum levels of the quantizer) is given by

$$Q_S = 2\alpha_i \sigma_i \quad (6-1)$$

where α_i is the amplitude loading factor of the i th quantizer. The step size of the quantizer is

$$E_0 = \frac{Q_S}{2^{n_i}} = \frac{2\alpha_i \sigma_i}{2^{n_i}} \quad (6-2)$$

If there is a large number of steps, then the mean-square quantizing noise error is one-twelfth of the square of the step size.⁴⁸ That is,

$$\frac{\overline{\epsilon^2}}{12} = \frac{E_0^2}{3} = \frac{1}{3} \frac{\alpha_i^2 \sigma_i^2}{2^{2n_i}} \quad (6-3)$$

When there is a finite number of steps, 2^{n_i} , the expression for the mean-square quantizing noise error is more involved, since the probability distribution of the coefficients should be considered. The probability distribution of the coefficients can be assumed to be Gaussian, since the coefficients are obtained from a linear combination of spatial domain samples:

$$P(x) = \frac{1}{\sqrt{2\pi} \sigma_i} \exp(-x^2/2\sigma_i^2) \quad (6-4)$$

The exact analytical expression for quantizing noise power is known for stationary Gaussian signals.⁴⁹ The quantizing noise power, N_Q , which is composed of two parts, granular noise power, N_G , and overload noise power, N_O , can be computed as follows:

$$\frac{N_G}{\sigma_i^2} = \frac{2}{3} \frac{\alpha_i^2}{2^{2n_i}} \operatorname{erf}(\alpha_i) \quad (6-5a)$$

$$\frac{N_O}{\sigma_i^2} = (1 + \alpha_i^2) [1 - 2 \operatorname{erf}(\alpha_i)] - \sqrt{\frac{2}{\pi}} \alpha_i e^{-\alpha_i^2/2} \quad (6-5b)$$

$$N_Q = N_G + N_O \quad (6-5c)$$

where $\operatorname{erf}(\alpha_i)$ is the error function.^{2 3}

For large α_i , equation (6-5a) yields equation (6-3). This term is what is generally known as quantizing noise because, when a suitable value of α_i is used, the overload noise is less than the granular noise and is therefore ignored.

Equation (6-5) can be used to compute N_Q/σ_i^2 and the signal-to-quantizing noise ratio:

$$(S/N_Q)_i = 10 \log_{10} \left(\frac{\sigma_i^2}{N_Q} \right)$$

for various n_i . Figure 6-1 shows the $(S/N_Q)_i$ values versus α_i for various n_i . The optimum α_i , which maximizes the signal-to-quantizing noise ratio determined from these curves, is given in Table 6-1.

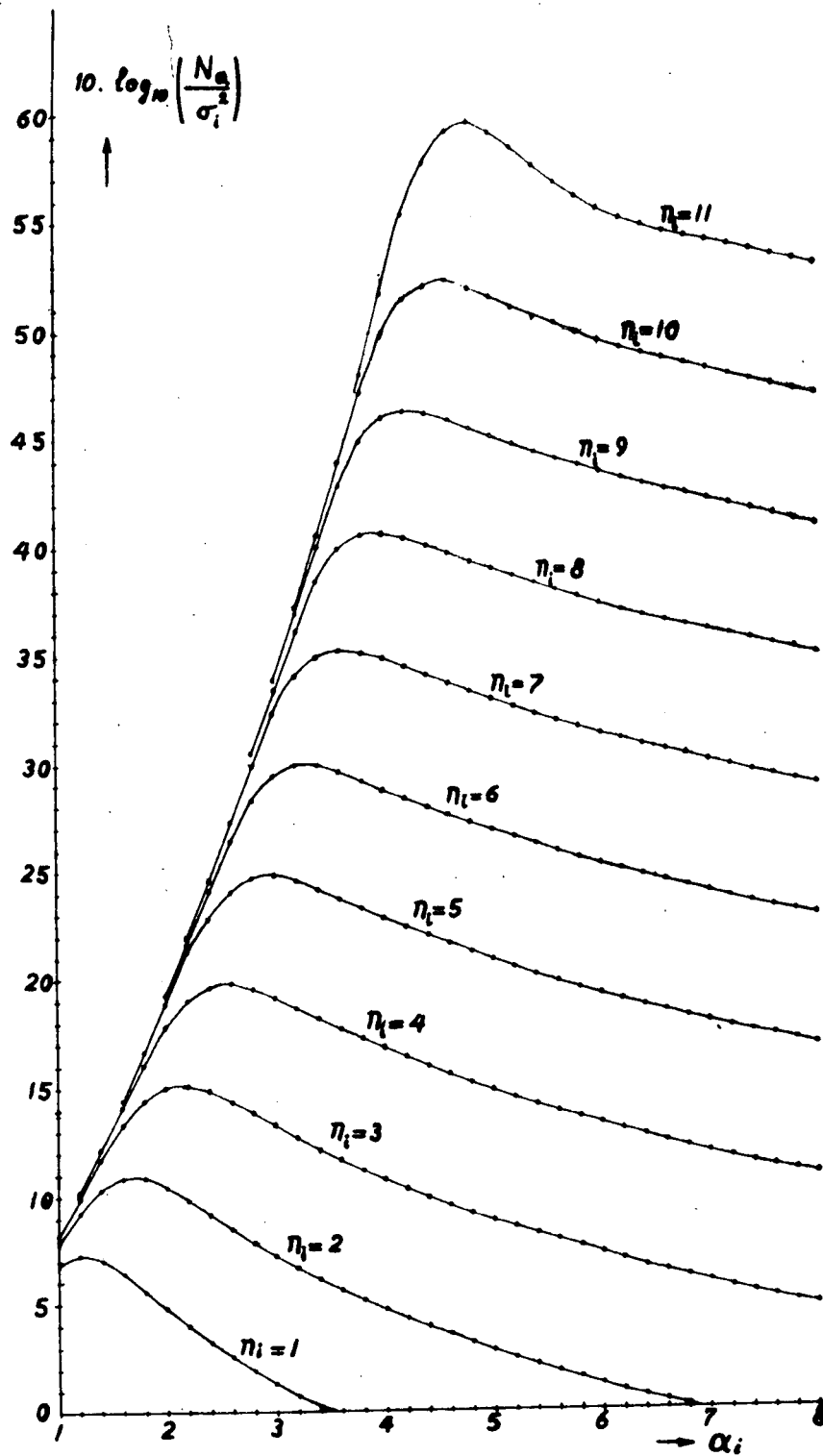


Figure 6-1. Signal-to-Quantizing Noise vs Amplitude Loading Factor, α , for Various Bit Assignments

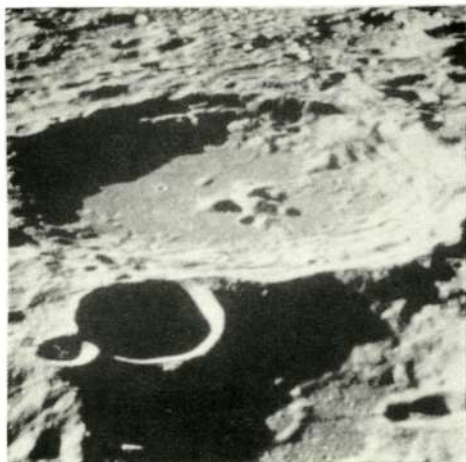
Table 6-1. Optimum Amplitude Loading Factor of the Quantizer in the Walsh-Hadamard Transform Domain

Number of Bits per Sample Assigned to the i th Coefficient (n_i)	Amplitude Loading Factor for the Uniform Quantization of the i th Coefficient (α_i)	
	Optimum Value	Value Used
1	1.25	1.5
2	1.8	2.0
3	2.1	2.5
4	2.5	3.0
5	3.0	3.4
6	3.3	3.8
7	3.7	4.1
8	4.0	4.4
9	4.3	4.7
10	4.5	4.9
11	4.7	5.1

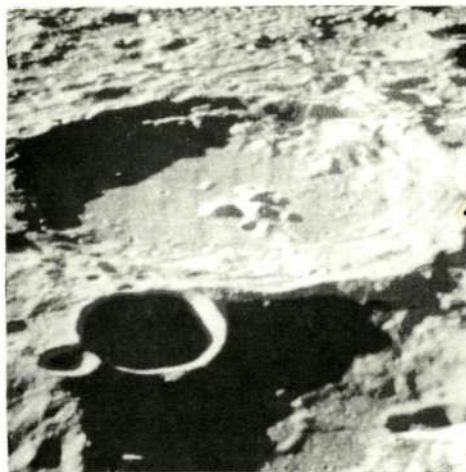
6.3 PCM IMAGE PROCESSING

The moonscape image has been processed by PCM at three different bit rates: seven, four, and two bits per sample. Figures 6-2a, 6-2b, and 6-2c are the resulting pictures for each of these cases. Uniform quantizing with a mid-riser was employed in each case, since this prevented the loss of one quantizing step. The quantizing step size was referenced to the full black-to-white range observed at the flying spot scanner output. For our computer, this corresponded to 2^{15} .

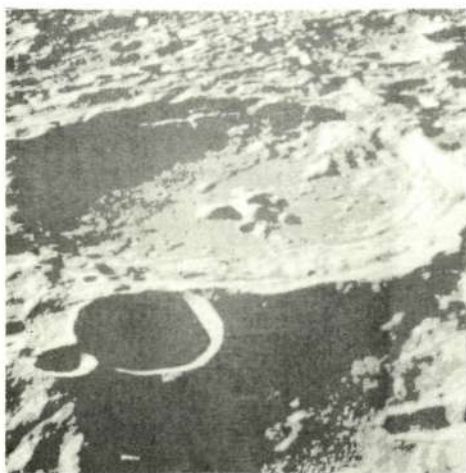
For each PCM bit rate, the error caused by quantization was determined. This was used as a basis for comparing the performance of the various transformations at the same average bit rates



a. PCM, M=7 bits/sample



b. PCM, M=4 bits/sample



c. PCM, M=2 bits/sample

Reproduced from
best available copy.

Figure 6-2. PCM Processed Images

in terms of reduction in quantization error. The quantizing noise was determined by computing the rms error difference between the source picture elements and the reconstructed picture elements.

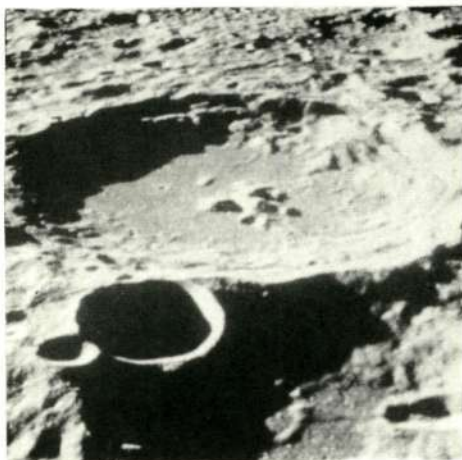
Figures 6-2a, 6-2b, and 6-2c are the pictures resulting from PCM processing at rates of seven, four, and two bits per sample, respectively. (A sample is equal to a picture element.) Processing at seven bits per sample (Figure 6-2a) produced a very high-quality picture. Subjectively, there was little difference between this picture and the original picture. The quantizing noise variance measured for this case was 5.43×10^3 . This is 53 dB below the peak picture element value of 2^{15} .

Figure 6-2b shows the result of PCM processing using four bits per sample. Contouring is becoming apparent (note the medium-sized crater). The quantizing noise variance in this case is 3.09×10^5 , which is 45 dB below the peak range value of 2^{15} .

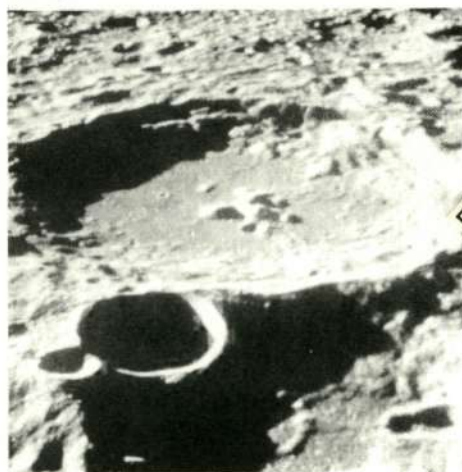
The PCM result for two bits per sample is given in Figure 6-2c. The quantizing noise variance was 5.90×10^6 , or 22.5 dB below the peak value of 2^{15} . Careful observation shows the loss of small details such as the smallest craters that were quite apparent in the seven- and four-bit-per-sample cases. There is also a "wash-out" effect because of the loss of the black and white extreme values. This could have been offset by rescaling to conform to the full black-to-white range of the flying spot scanner; however, doing so would not have improved the resolution.

6.4 FOURIER TRANSFORM IMAGE PROCESSING

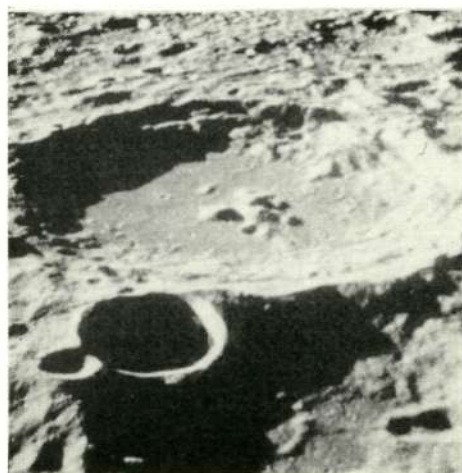
Fourier processing of the moonscape image at an average of seven, four, and two bits per sample is shown in Figure 6-3. For the seven-bit-per-sample processing, the coefficient bit assignments shown in Table 5-1 were used. The amplitude loading factor was $\alpha = 8$ for all coefficients. The quantizing noise was



a. DFT, Average 7 bits/sample, $\alpha=8$



b. DFT, Average 4 bits/sample, $\alpha=4$



c. DFT, Average 2 bits/sample, $\alpha=4$

Reproduced from
best available copy.

Figure 6-3. 2D-DFT Processed Images

56.9 dB below the peak value of 2^{15} , or 3.1 dB less than the value computed for seven-bit PCM processing. Subjectively, the picture enhancement due to the decreased quantizing noise is less readily apparent than in the seven-bit PCM picture.

Next, Fourier processing was performed at an average of four bits per sample. The bit assignments used are listed in Table 5-2. Two values of the amplitude loading factor, $\alpha = 4$ and $\alpha = 8$, were examined. The quantizing noise variances were 39 dB for $\alpha = 4$ and 37.4 dB for $\alpha = 8$. Figure 6-3b is the picture produced for $\alpha = 4$. When compared with the picture produced by four-bit-per-sample PCM processing, this picture shows considerable improvement, especially with regard to the elimination of contouring.

Fourier processing was also performed at two bits per sample using the bit assignment given in Table 5-3 and for $\alpha = 2, 4, 6$, and 8. These values resulted in quantizing noise variances of 27.5, 32.3, 28.6, and 25.8 dB, respectively, below the peak value of 2^{15} . The best result, shown in Figure 6-3c, was obtained for $\alpha = 4$ and is 9.8 dB less than the quantizing obtained for PCM processing at the two-bit-per-sample rate. When compared with the result of two-bit-per-sample PCM processing shown in Figure 6-2, Figure 6-3c reveals significantly improved resolution of the fine detail and no contouring. However, a block structure, which may possibly be removed by using slightly different processing strategy (such as overlapping of processing blocks), is apparent. It will be seen later that this block structure is less apparent in the Hadamard case and almost disappears in the Karhunen-Loève case.

6.5 HADAMARD TRANSFORM IMAGE PROCESSING

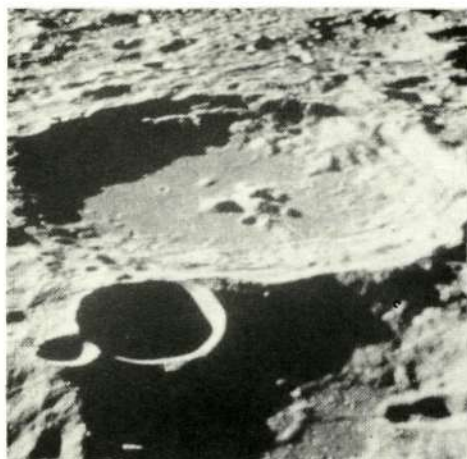
Hadamard processing using average bit rates of seven, four, and two bits per sample has been performed by using the same moonscape picture processed by the PCM and Fourier methods. The method used was presented in Section 5.5. The results of the Hadamard image processing are shown in Figure 6-4.

Figure 6-4a was obtained for an average rate of seven bits per sample. The bit distribution used is shown in Table 5-4 and the value of α was eight. The computed variance in this case was 54.6 dB below the peak of 2^{15} , which is 1.6 dB less variance than was observed for PCM processing. As is seen, the processed picture possesses excellent quality.

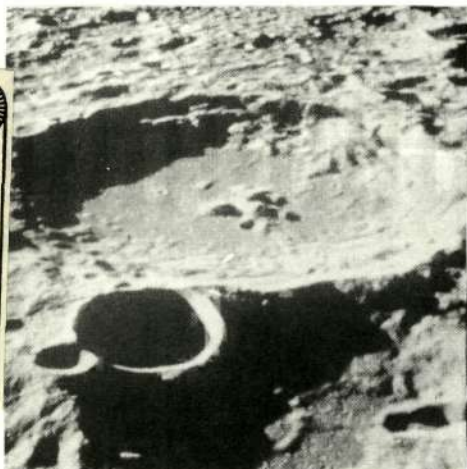
Figure 6-4b is the picture obtained at an average of four bits per sample. In this case, the computed quantizing noise variance was 38.1 dB below the peak value of 2^{15} , which is 3.1 dB less than that observed for the PCM case. When compared with the PCM case of Figure 6-2, this picture shows that contouring has been eliminated. When it is compared with the picture for Fourier processing, it can be seen that there is little perceptible difference between the two.

An alternative method of Hadamard processing at an average rate of four bits per sample was also investigated. This consisted of assigning new optimum values of α (given in Table 6-1) for each of the bit assignments. This procedure resulted in a quantizing noise variance that was 39.1 dB below the peak value of 2^{15} , 4.1 dB less than the PCM processing variance, and 1 dB less than the variance obtained when a value of $\alpha = 4$ was used for all of the bit assignments for four-bit-per-sample Hadamard processing.

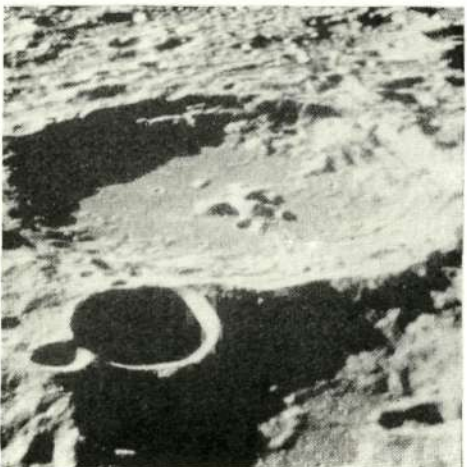
Figure 6-4c is for Hadamard processing at an average rate of two bits per sample. The calculated quantizing noise variance for $\alpha = 4$ was found to be 30.4 dB below the peak value of



a. Hadamard, Average 7 bits/sample, $\alpha=8$



b. Hadamard, Average 4 bits/sample, $\alpha=4$



c. Hadamard, Average 2 bits/sample, $\alpha=4$

Reproduced from
best available copy.

Figure 6-4. 2D-DWT (Hadamard) Processed Images

2^{15} . This is 7.9 dB less quantizing noise variance than was calculated for PCM processing. Examination of the picture shows that most of the fine detail which was lost in two-bit-per-sample PCM processing was retained. Comparison with Figure 6-3 shows about the same amount of resolution for both the Hadamard and Fourier methods. As in the Fourier case, some block structure is apparent. Also, there is a noticeable amount of snow-like noise.

6.6

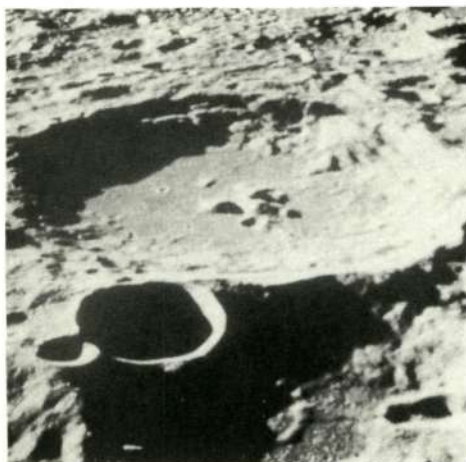
KARHUNEN-LOÈVE TRANSFORM IMAGE PROCESSING

Karhunen-Loève processing using average bit rates of seven, four, and two bits per sample has been performed using the same mooncape picture processed by using the PCM, Fourier, and Hadamard methods. The method used was presented in Section 5.6. The Karhunen-Loève transformed images are shown in Figure 6-5.

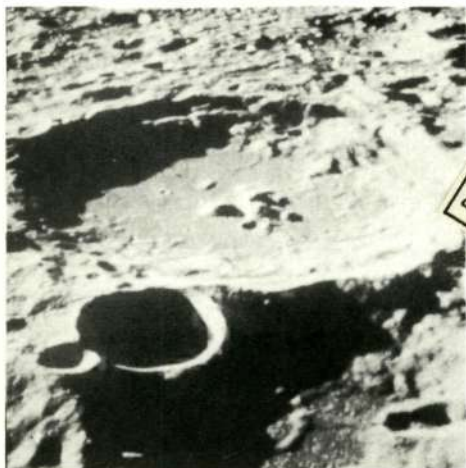
Figure 6-5a was obtained for an average bit rate of seven bits per sample. The bit distribution used is shown in Table 5-10, and the value of α was 8. The quantizing noise variance was 62.9 dB below the peak value of 2^{15} , which is better than that produced by any of the other processing methods and 9.9 dB less than that computed for PCM processing. The picture quality, as should be expected, was excellent.

Figure 6-5b was obtained for an average of four bits per sample and $\alpha = 4$. Its computed quantizing noise variance was 39.8 dB below the peak value of 2^{15} , which is 4.8 dB better than for the PCM case. It is believed that some additional improvement may be achieved by varying α . The resulting picture is similar in appearance to those obtained for the Fourier and Hadamard transformations.

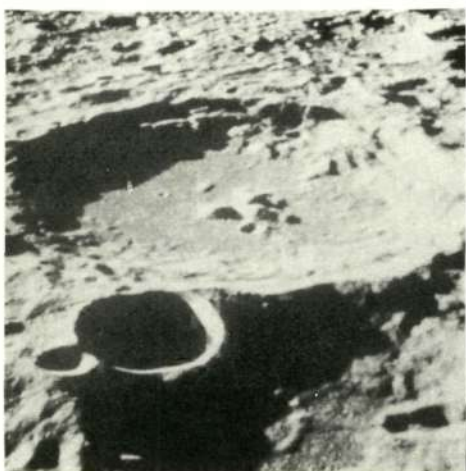
Figure 6-5c is for Karhunen-Loève processing using an average rate of two bits per sample and $\alpha = 4$. The bit



a. KLT, Average 7 bits/sample, $\alpha=4$



b. KLT, Average 4 bits/sample, $\alpha=4$



c. KLT, Average 2 bits/sample, $\alpha=4$

Reproduced from
best available copy.

Figure 6-5. 2D-DKLT Processed Images

distribution used for this case is shown in Table 5-12. The calculated quantizing noise variance for $\alpha = 4$ was found to be 36.2 dB below the peak value of 2^{15} . This quantizing noise is 13.7 dB less than that produced by two-bit-per-sample PCM processing and is equivalent to that produced by four-bit PCM processing.

As will be seen by comparing the processed pictures, however, the two-bit-per-sample Karhunen-Loève transform does not produce the perceptible contouring which occurs as a result of four-bit-per-sample PCM processing. The reason for this is quite apparent. In PCM processing, there is no interpolation among the image samples; i.e., each is presented exclusively. The boundaries along which quantizing thresholds are crossed become perceptible even for relatively small quantizing step sizes. On the other hand, for the transform processed picture, each picture element appearing in the reconstructed image is the weighted average of N two-dimensional eigenfunctions. This results in interpolation of the picture element values in the reconstructed image and virtually removes contouring at low bit rates. It should be further noted that the block structure that was apparent in the Fourier case is almost totally removed in the Karhunen-Loève case. Also, there is much less snow-type noise than was observed in the other cases.

The eigenfunctions used for processing were identical in the horizontal and vertical directions. Actually, the correlations in the x and y directions differed by a small amount, as shown in Figure 5-4. Processing using nonidentical eigenfunctions (derived from the preceding correlation functions) in the x and y directions did not significantly enhance the reconstructed picture and is therefore not shown.

6.7 SUMMARY OF ORTHOGONAL TRANSFORM TECHNIQUES FOR IMAGE PROCESSING

Image processing using PCM, 2D-DFT, 2D-DHT, and 2D-KLT was performed on a moonscape slide provided by NASA/MSL at bit rates averaging seven, four, and two bits per sample. The image was resolved into 512 x 512 picture elements using a flying spot scanner. For transformation processing, this array was further divided into 1024 16 x 16 element subarrays. The pictures resulting from the various processing methods were shown in Figures 6-2, 6-3, 6-4, and 6-5.

To quantitatively assess the results obtained, the rms difference between preprocessed and postprocessed values of array elements was computed. This was designated as the quantizing noise variance. PCM processing was used as a basis for comparison of quantizing noise variance and, for each transformation processing, the reduction in quantizing noise (or improvement in signal-to-quantizing noise ratio) achieved for each of the bit rates was determined.

The results obtained for Fourier processing (2D-DFT) are summarized in Table 6-2. Note that, in some cases, the value of the amplitude loading was modified to experimentally determine an optimum value. Thus, at an average rate of four bits per sample, an amplitude loading factor of 4 produced a better result than a value of 8, and, for a rate of two bits per sample, a value of 4 also produced the best result.

Table 6-3 is a comparison of the results for all of the processing methods using near-optimum amplitude loading factors. This table shows that the amount of improvement in S/N_Q increases at low bit rates for each transformation method. It also shows that the Karhunen-Loève transformation produces the best results, the Fourier transformation produces the next best results, and the Hadamard transformation produces the worst results. All of the transformations produce better results than PCM.

Table 6-2. Signal-to-Quantizing Noise Ratio Improvement (Λ_Q) of 2D-DFT Over PCM as Applied to Moonscape Slide

Average Number of Bits per Sample Used in the Transform Domain	Amplitude Loading Factor (α) of the Quantizer in the Transform Domain	Improvement in S/N_Q (Λ_Q) Over PCM
2	2	5.0
	4	9.8
	6	6.1
	8	3.3
4	4	4.0
	8	2.4
7	8	3.9

Table 6-3. Signal-to-Quantizing Noise Ratio Improvement of Two-Dimensional Orthogonal Transforms Over PCM as Applied to Moonscape Slide

Average Number of Bits per Sample Used in the Transform Domain, \bar{n}_i	Amplitude Loading Factor of the Quantizer in the Transform Domain, α_i	Improvement in S/N_Q Over PCM (dB)		
		Fourier	Hadamard	Karhunen-Loève
2	4	9.8	7.9	13.7
4	4	4.0	3.1	4.8
7	8	3.9	1.6	9.9

A very important property of transform processing that is not revealed in the S/N_Q improvements is the subjective effect of contouring. It is basically true of all of the transformation processing methods that, at low bit rates, the effect of contouring is virtually eliminated. This is because each image element in the reconstructed transformed image is the weighted average of N

two-dimensional orthogonal functions; the interpolation provided thereby eliminates the contouring that would occur in PCM processing at the same bit rate.

Because of the rather good results observed in this study, the use of orthogonal transformation processing of images should be studied further. The Karhunen-Loève transformation is very effective, but its implementation is the most complex. However, this should not eliminate it from serious consideration since the cost of digital processing methods is rapidly decreasing. The Hadamard transform method is the simplest to implement and certainly warrants continued effort, especially since it significantly reduces contouring at low bit rates.

7. BIBLIOGRAPHY

7.1 RATE-DISTORTION CRITERIA

1. Berger, T. "Information Rates of Wiener Processes," IEEE Transactions on Information Theory, Vol. IT-16, March 1970, pp. 134-139.
2. Bunin, B. J. "Rate Distortion Functions For Gaussian Markov Processes," Bell System Technical Journal, November 1969, pp. 3059-3074.
3. Davisson, L. D. "Rate Distortion Theory and Data Compression," Proceedings of the Princeton University Conference, Princeton, New Jersey, March 1970.
4. Gallagher, R. G. Information Theory and Reliable Communication, New York: John Wiley, 1969.
5. Goblick, T. J., Jr. "Theoretical Limitations on the Transmission of Data from Analog Sources," IEEE Transactions on Information Theory, Vol. IT-11, No. 4, October 1965.
6. Goblick, T. J. and Holsinger, J. L. "Analog Source Digitization: A Comparison of Theory and Practice," IEEE Transactions on Information Theory, Vol. IT-13, No. 2, April 1967, pp. 323-326.
7. Gray, R. M. "Information Rates of Autoregressive Processes," IEEE Transactions on Information Theory, Vol. IT-16, July 1970, pp. 412-421.

8. Haskell, B. G. "The Computation and Bounding of Rate-Distortion Functions," IEEE Transactions on Information Theory, Vol. IT-15, September 1969, pp. 525-531.
9. Hayes, J. F. "Rate Distortion Bound for Single Sample Encoding," Report No. TR-EE69-2, School of Electrical Engineering, Purdue University, Lafayette, Indiana, January 1969.
10. Hayes, J. F., Habibi, A., and Wintz, P. A. "Rate Distortion Function for a Gaussian Model of Image," IEEE Transactions on Information Theory (Correspondence), Vol. IT-16, July 1970, pp. 507-509.
11. Kellogg, W. C. "Information Rates in Sampling and Quantization," IEEE Transactions on Information Theory, Vol. IT-13, July 1967, pp. 506-511.
12. Kolmogorov, A. N. "On the Shannon Theory of Information in the Case of Continuous Signals," IRE Transactions on Information Theory, Vol. IT-2, December 1956, pp. 102-108.
13. Lynch, T. J. "Data Compression with Error-Control Coding for Space Telemetry," NASA Technical Report NASA TR R-261, Goddard Space Flight Center, Greenbelt, Maryland, June 1967.
14. McDonald, R. A. and P. M. Schultheiss. "Information Rates of Gaussian Signals Under Criteria Constraining the Error Spectrum," Proceedings of the IEEE, Vol. 52, April 1964, pp. 415-416.
15. Pinkston, J. T. "Encoding Independent Sample Information Sources," Technical Report No. 462, M.I.T. Research Laboratory of Electronics, Cambridge, Massachusetts, October 31, 1967.

16. Sakrison, D. J. "A Geometric Treatment of the Source Encoding of a Gaussian Random Variable," IEEE Transactions on Information Theory, Vol. IT-14, May 1968, pp. 481-486.
17. Sakrison, D. J. "The Rate Distortion Function of a Class of Sources," Information and Control, Vol. 15, No. 2, August 1969, pp. 165-195.
18. Sakrison, D. J. "The Rate Distortion Function of a Gaussian Process with a Weighted Square Error Criterion," IEEE Transactions on Information Theory (Correspondence), Vol. IT-14, May 1968, pp. 506-508.
19. Sakrison, D. J. "The Rate of a Class of Random Functions," IEEE Transactions on Information Theory, Vol. IT-16, January 1970, pp. 10-16.
20. Shannon, C. E. "A Mathematical Theory of Communications," Bell System Technical Journal, Vol. 27, July 1948, pp. 379-423; October 1948, pp. 623-656.
21. Shannon, C. E. "Coding Theorems for Discrete Sources With a Fidelity Criterion," Information and Decision Processes, ed. by Robert E. Machol, New York: McGraw-Hill, 1960, pp. 93-126.

7.2 DISCRETE FOURIER TRANSFORM

1. Bergland, G. D. "A Radix-Eight Fast Fourier Transform Subroutine for Real-Valued Series," IEEE Transactions on Audio and Electroacoustics, Vol. AU-17, No. 2, June 1969, pp. 138-144.

2. Cooley, J. W. and J. W. Tukey. "An Algorithm for the Machine Calculation of Complex Fourier Series," Mathematics of Computation, Vol. 19, April 1965, pp. 297-301.
3. G-AE Subcommittee on Measurement Concepts, "What is the Fast Fourier Transform," IEEE Transactions on Audio and Electroacoustics, Vol. AU-15, June 1967, pp. 45-55; Proceedings of the IEEE, Vol. 55, October 1967, pp. 1664-1674.
4. Gentleman, W. M. and G. Sande. "Fast Fourier Transforms for Fun and Profit," AFIPS Proceedings (1966 Fall Joint Computer Conference), Spartan Books, 1966, Vol. 29, pp. 563-578.
5. Singleton, R. C. "A Short Bibliography on the Fast Fourier Transform," IEEE Transactions on Audio and Electroacoustics, Vol. AU-17, No. 2, June 1969, pp. 166-169.
6. Singleton, R. C. "An Algorithm for Computing the Mixed Radix Fast Fourier Transform," IEEE Transactions on Audio and Electroacoustics, Vol. AU-17, No. 2, June 1969, pp. 93-103.
7. Stockham, T. G. "High Speed Convolution and Correlation," AFIPS Proceedings (Spring Joint Computer Conference), Spartan Books, 1966, Vol. 28, pp. 229-233.

7.3 DISCRETE WALSH-HADAMARD TRANSFORM

1. Golomb, S. W. et al., Digital Communications, New Jersey: Prentice Hall, 1964.
2. Harmuth, H. F. Transmission of Information by Orthogonal Functions, New York: Springer-Verlag, 1969.

3. Pease, M. C. "An Adaptation of the Fast Fourier Transform for Parallel Processing," Journal of the Association for Computing Machinery, Vol. 15, No. 2, April 1968, pp. 252-264.
4. Pratt, W. K., J. Kane, and H. C. Andrews. "Hadamard Transform Image Coding," Proceedings of the IEEE, Vol. 57, January 1969, pp. 58-68.
5. Robinson, G. S., C. F. Chao, and R. L. Granger. "Speech Compression by Linear Transformations," COMSAT Laboratories Technical Memorandum CL-9-70, March 9, 1970.
6. Walsh, J. L. "A Closed Set of Orthogonal Functions," American Journal of Mathematics, Vol. 55, 1923, pp. 5-24.
7. Welch, J. E., Jr., and D. F. Guinn. "The Fast Fourier-Hadamard Transformation and Its Use in Signal Representation and Classifications," EASCON Record [1968 Electronics and Aerospace Systems Convention], pp. 561-571.

7.4 DISCRETE HAAR TRANSFORM

1. Andrews, H. C. and K. L. Caspari. "A Generalized Technique for Spectral Analysis," IEEE Transactions on Computers, Vol. C-19, No. 1, January 1970, pp. 16-25.
2. Bremerman, H. J. "Pattern Recognition, Functionals and Entropy," IEEE Transactions on Bio-Medical Engineering, Vol. BME-15, No. 3, July 1968, pp. 201-207.
3. Collatz, L. Functional Analysis and Numerical Mathematics, New York: Academic Press, 1962.

4. Gelbaum, B. R. "On the Functions of Haar," Annals of Mathematics, Vol. 51, No. 1, January 1950, pp. 26-36.
5. Haar, A. "Zur Theorie der Orthogonalen Funktionensysteme," Mathematische Annalen, Vol. 69, 1910, pp. 331-371.
6. Robinson, G. S., C. F. Chao, and R. L. Granger. "Speech Compression by Linear Transformations," COMSAT Laboratories Technical Memorandum CL-9-70, March 9, 1970.
7. Sitton, G. A., "Acoustic Segmentation of Speech," Rice University Ms Thesis, June 1969.
8. Walsh, J. L. "A Closed Set of Orthogonal Functions," American Journal of Mathematics, Vol. 55, 1923, pp. 5-24.
9. Watari, C. "A Generalization of Haar Functions," Tohoku Mathematical Journal, Vol. 8, 1956, pp. 286-290.

7.5 KARHUNEN-LOÈVE TRANSFORM

1. Andrews, H. C. and W. K. Pratt. "Transform Image Coding," Proceedings of the Symposium on Computer Processing in Communications, Polytechnic Institute of Brooklyn, April 8-10, 1969, pp. 63-83.
2. Davenport, W. B. and W. L. Root. Random Signals and Noise, New York: McGraw-Hill, 1958.

3. Fukunaga, K. and W. L. G. Koontz. "Application of the Karhunen-Loeve Expansion to Feature Selection and Ordering," IEEE Transactions on Computers, Vol. C-19, No. 4, April 1970, pp. 311-318.
4. Fukunaga, K. and W. L. G. Koontz. "Representation of Random Processes Using the Finite Karhunen-Loève Expansion," Information and Control, Vol. 16, 1970, pp. 85-101.
5. Goodman, L. M. "A Binary Linear Transformation for Redundancy Reduction," Proceedings of the IEEE (Letters), Vol. 55, No. 3, March 1967, pp. 467-468.
6. Greenstadt, J. "The Determination of the Characteristic Roots of a Matrix by the Jacobi Method," Mathematical Methods for Digital Computers, ed. by A. Ralston and H. S. Wilf, New York: John Wiley, 1960, pp. 84-91.
7. Habibi, A. and P. Wintz. "Optimum Linear Transformations for Encoding 2-Dimensional Data," Technical Report No. TR-EE69-15, Purdue University, May 1969.
8. Huang, J. J. Y. and P. M. Schultheiss. "Block Quantization of Correlated Gaussian Random Variables," IEEE Transactions on Communications Systems, Vol. CS-11, No. 3, September 1963, pp. 289-296.
9. Kramer, J. P. and M. V. Mathews. "A Linear Coding for Transmitting a Set of Correlated Signals," IRE Transactions on Information Theory, Vol. IT-2, September 1956, pp. 41-46.

10. Ortega, J. "The Givens-Householder Method for Symmetric Matrices," Mathematical Methods for Digital Computers, ed. by A. Ralston and H. S. Wilf, New York: John Wiley, 1960, pp. 84-91.
11. Papoulis, A. Probability, Random Variables and Stochastic Processes, New York: McGraw-Hill, 1965.
12. Watanabe, S. "Karhunen-Loève Expansion and Factor Analysis," Transactions of the Fourth Prague Conference on Information Theory, 1965, pp. 635-660.
13. Whelchel, J. E., Jr. and D. F. Guinn. "The Fast Fourier-Hadamard Transform and Its Use in Signal Representation and Classification," EASCON Record [1968 Electronics and Aerospace Systems Convention], pp. 561-573.
14. Young, T. Y. and W. F. Huggins. "On the Representation of Electrocardiograms," IEEE Transactions on Bio-Medical Electronics, July 1963.

7.6 IMAGE PROCESSING BY LINEAR TRANSFORMATIONS

1. Agarwal, V. K. and T. J. Stephens, Jr. "On-Board Data Processor-Picture Bandwidth Compression," IRAD Report No. R-7333.3-32, TRW System Group, California, February 1970.
2. Anderson, G. B. and T. S. Huang. "Picture Bandwidth Compression by Piecewise Fourier Transformation," Proceedings of the Purdue Centennial Year Symposium on Information Processing, 1969.

3. Anderson, G. B. and T. S. Huang. "Piecewise Fourier Transformation for Picture Bandwidth Compression," IEEE Transactions on Communications Technology, Vol. COM-19, No. 2, April 1971.
4. Andrews, H. C. "Fourier and Hadamard Image Transform Channel Error Tolerance," Proceedings of the UMR-Mervin J. Kelly Communications Conference, University of Missouri, Rolla, Missouri, October 1970, Paper No. 10-4.
5. Andrews, H. C. and W. K. Pratt. "Digital Image Transform Processing," Proceedings of the 1970 Symposium on the Applications of Walsh Functions, Washington, D. C., April 1970, pp. 183-194.
6. Andrews, H. C. and W. K. Pratt. "Fourier Transform Coding of Images," presented at the Hawaii International Conference on System Sciences, January 1968.
7. Andrews, H. C. and W. K. Pratt. "Television Bandwidth Reduction by Encoding Spatial Frequencies," Journal of the Society of Motion Picture and Television Engineers, Vol. 77, December 1968, pp. 1279-1281.
8. Bowyer, D. "Walsh Functions, Hadamard Matrices and Data Compression," Proceedings of the 1971 Symposium on the Applications of Walsh Functions, Washington, D. C., April 1971.
9. Claire, E. J., S. M. Farber, and R. R. Green. "Practical Techniques for Transform Data Compression," Proceedings of the 1971 Symposium on the Applications of Walsh Functions, Washington, D. C., April 1971.

10. Enomoto, H. and K. Shibata. "Orthogonal Transform Coding System for Television Signal," Television, Vol. 24, No. 2, February 1970 (Japan), pp. 98-108.
11. Enomoto, H. and K. Shibata. "Orthogonal Transform Coding System for Television Signals," Proceedings of the 1971 Symposium on the Applications of Walsh Functions, Washington, D. C., April 1971.
12. Fino, B. J. "Traitement d'images pour Transformations Orthogonales," CNET Report, ITD/TTI, September 1971.
13. Habibi, A. and P. A. Wintz. "Image Coding by Linear Transformation and Block Quantization," IEEE Transactions on Communications Technology, Vol. COM-19, No. 1, February 1971, pp. 50-62.
14. Habibi, A. and P. A. Wintz. "Linear Transformations for Encoding 2-Dimensional Sources," Technical Report No. TR-EE70-2, School of Electrical Engineering, Purdue University, Lafayette, Indiana, June 1970.
15. Habibi, A. and P. A. Wintz. "Optimum Linear Transformations for Encoding Two-Dimensional Data," Technical Report No. TR-EE69-15, School of Electrical Engineering, Purdue University, Lafayette, Indiana, May 1969.
16. Hatori, M. and Y. Taki. "Digital Communication Using Hadamard Transformation," Electronics & Communications in Japan, Vol. 52-B, No. 5, 1969, pp. 79-87.

17. Huang, T. S. and J. W. Woods. "Picture Bandwidth Compression by Block Quantization," presented at the 1969 International Symposium on Information Theory, Ellenville, New York.
18. Kennedy, J. D. "Walsh Function Imagery Analysis," Proceedings of the 1971 Symposium on the Applications of Walsh Functions, Washington, D. C., April 1971.
19. Kennedy, J. D., S. J. Clark, and W. A. Parkyn. "Digital Imagery Data Compression Techniques," Report MDC G0402, McDonnell Douglas Astronautics Company, California, January 1970.
20. Landau, H. J. and D. Slepian. "Some Computer Experiments in Picture Processing for Bandwidth Reduction," Bell System Technical Journal, May-June 1971, pp. 1525-1540.
21. Marano, P. H. and P. Y. Schwartz. "Compression d'Information sur la transformee de Fourier d'une image," L'onde Electrique, Vol. 50, Fasc. 11, December 1970, pp. 908-919.
22. Nitadori, K. "Linear Transformation Coding and Predictive Coding - Two Methods of Digital Encoding for Continuous Sources with Discrete Parameters," Electronics & Communications in Japan, Vol. 53-A, No. 2, 1970, pp. 37-45.
23. Ponchin, J. "Utilisation de la Transformation de Hadamard pour le Codage et la Compression de Sigaux d'images, CNET Report, ITD/TTI/49, 1971.
24. Pratt, W. K. "Linear and Nonlinear Filtering in the Walsh Domain," Proceedings of the 1971 Symposium on the Applications of Walsh Functions, Washington, D. C., April 1971.

25. Pratt, W. K. and H. C. Andrews, "Application of Fourier-Hadamard Transformation to Bandwidth Compression," presented at the 1969 Symposium on Picture Bandwidth Compression, M.I.T., Cambridge, Massachusetts.
26. Pratt, W. K. and H. C. Andrews. "Transform Image Coding," USCEE Report No. 387, University of California, March 1970.
27. Pratt, W. K., J. Kane, and H. C. Andrews. "Hadamard Transform Image Coding," Proceedings of the IEEE, Vol. 57, January 1969, pp. 58-68.
28. Sakrison, D. J. and R. Algazi. "A Comparison of Line-by-Line and Two-Dimensional Encoding of Random Images," IEEE Transactions on Information Theory, Vol. IT-17, July 1970, pp. 386-398.
29. Schwartz, P. Y. "Analyse de la Compression d'information sur la transformee de Fourier d'une image," Annales des Telecommunications, Mars-Avril 1971.
30. Schwartz, P. Y. "Analysis of Fourier Compression of Images," CNET Note Technique Interne EST/TDS/423, presented at the Symposium on Picture Coding, Raleigh, North Carolina, September 1970.
31. Shibata, K., T. Ohira, and S. Terauchi. "Color Television Signal Orthogonal Transform PCM Terminal Experimental Equipment," IEECJ Transactions on Communications Systems, Vol. CS70-47, July 29, 1970 (Japan).
32. Woods, J. W. "Video Bandwidth Compression by Linear Transformation," Quarterly Progress Report No. 91, M.I.T. Research Laboratory of Electronics, October 15, 1968, pp. 219-224.

7.7 PICTURE TRANSMISSION BY PULSE-CODE MODULATION

1. Braidy, R., B. Laurent, and B. Picot. "Systeme de Television pour Transmission P.C.M.," L'onde Electrique, Vol. 49, Fasc. 8, September 1969, pp. 874-884.
2. Bruce, R. A. "Optimum Pre-emphasis and De-emphasis Networks for Transmission of Television by PCM," IEEE Transactions on Communications Technology, Vol. COM-12, September 1964, pp. 91-96.
3. Carbrey, W. M. "Video Transmission over Telephone Cable Pairs by Pulse Code Modulation," Proceedings of the IRE, Vol. 48, September 1960, pp. 1546-1561.
4. Goodall, W. M. "Television by Pulse Code Modulation," Bell System Technical Journal, Vol. 1, January 1951, pp. 33-49.
5. Huang, T. S. "PCM Picture Transmission," IEEE Spectrum, Vol. 2, December 1965, pp. 57-63.
6. Huang, T. S. and Chikhaoui, "The Effect of BSC on PCM Picture Quality," IEEE Transactions on Information Theory, Vol. IT-13, April 1967, pp. 270-273.
7. Huang, T. S. and O. J. Tretiak, "Research in Picture Processing," Optical and Electrooptical Information Processing, ed. by T. Tippet, Cambridge, Massachusetts: M.I.T. Press, 1965, Chapter 3.
8. Huang, T. S., O. J. Tretiak, B. Prasada, and Y. Yamaguchi. "Design Considerations in PCM Transmission of Low-Resolution Monochrome Still Pictures," Proceedings of the IEEE, Vol. 55, No. 3, March 1967, pp. 331-335.

9. Kimme, E. G. and F. F. Kuo. "Synthesis of Optimal Filters for a Feedback Quantization System," IEEE Transactions on Circuit Theory, Vol. CT-10, September 1963, pp. 404-413.
10. Max, J. "Quantizing for Minimum Distortion," IRE Transactions on Information Theory, Vol. IT-16, March 1960, pp. 7-12.
11. Panter, P. F. and W. Dite. "Quantization Distortion in Pulse Count Modulation with Nonuniform Spacing of Levels," Proceedings of the IRE, Vol. 39, January 1951, pp. 44-48.
12. "Pulse Code Modulation for 625-Line Television," Electrical Communication, Vol. 41, No. 1, 1966, pp. 110-112.
13. Roberts, L. G. "Picture Coding Using Pseudo-Random Noise," IRE Transactions on Information Theory, Vol. IT-8, February 1962, pp. 145-154.
14. Schreiber, W. F. "Picture Coding," Proceedings of the IEEE, Vol. 55, No. 3, March 1967, pp. 320-330.
15. Scoville, F. W. and T. S. Huang. "The Subjective Effect of Spatial and Brightness Quantization in PCM Picture Transmission," Proceedings of the NEREM Convention Record, November 1965, pp. 234-235.
16. Spang, H. A. and P. M. Schultheiss. "Reduction of Quantizing Noise by Use of Feedback," IRE Transactions on Communications Systems, December 1962, pp. 373-380.
17. Weston, J. D. "Transmission of Television by Pulse Code Modulation," Electrical Communication, Vol. 42, No. 2, 1967, pp. 165-173.

7.8 KARHUNEN-LOÈVE TRANSFORM OF IMAGES

1. Davenport, W. B. and W. L. Root. An Introduction to the Theory of Random Signals and Noise, New York: McGraw-Hill, 1958.
2. Franks, L. E. "A Model for the Random Video Television Signals," Bell System Technical Journal, Vol. 45, April 1966, pp. 609-630.
3. Habibi, A. and P. Wintz. "Image Coding by Linear Transformation and Block Quantization," IEEE Transactions on Communications Technology, Vol. COM-19, No. 1, February 1971, pp. 50-62.
4. Kretzmer, E. R. "Statistics of Television Signals," Bell System Technical Journal, Vol. 31, July 1952, pp. 761-763.
5. Pratt, W. K. and H. C. Andrews. "Transform Image Coding," USCEE Report No. 387, University of California, March 1970.
6. Ray, W. D. and R. M. Driver. "Further Decomposition of the Karhunen-Loève Series Representation of a Stationary Random Process," IEEE Transactions on Information Theory, Vol. IT-16, No. 6, November 1970, pp. 663-668.
7. Van Trees, H. L. Detection, Estimation and Modulation Theory, Part I, New York: John Wiley, 1968.

7.9 QUANTIZATION EFFECTS IN IMAGE PROCESSING

1. Abate, J. E. "Linear and Adaptive Delta Modulation," Proceedings of the IEEE, Vol. 55, No. 3, March 1967, pp. 298-308.

2. Bennett, W. R. "Spectra of Quantized Signals," Bell System Technical Journal, Vol. 27, July 1968, pp. 446-472.
3. Papoulis, A. Probability, Random Variables and Stochastic Processes, New York: McGraw-Hill.

8. REFERENCES

1. J. P. Kramer and M. V. Mathews, "A Linear Coding for Transmitting a Set of Correlated Signals," IRE Transactions on Information Theory, Vol. IT-2, September 1956, pp. 41-46.
2. V. I. Kulya, "Experimental Investigation of the Correlation Relations in the Speech Spectrum and a Comparison of Some Variants of Orthogonal Vocoder," presented at the All-Union Conference on Coding Theory and Its Applications, Odessa, May 1963.
3. H. R. Crowther and C. M. Rader, "Efficient Coding of Vocoder Channel Signals using Linear Transformation," Proceedings of the IEEE (Letters), Vol. 54, November 1966, pp. 1594-1595.
4. J. J. Y. Huang and P. M. Schultheiss, "Block Quantization of Correlated Gaussian Random Variables," IEEE Transactions on Communications Systems, Vol. CS-11, No. 3, September 1963, pp. 289-296.
5. K. Nitadori, "Linear Transformation Coding and Predictive Coding - Two Methods of Digital Encoding for Continuous Sources with Discrete Parameters," Electronics & Communications in Japan, Vol. 53-A, No. 2, 1970, pp. 37-45.
6. R. G. Gallager, Information Theory and Reliable Communication, New York: John Wiley, Chapter 9, 1969.
7. D. H. Hampsher, Communications System Engineering Handbook, New York: McGraw-Hill, Chapter 3, 1967.

8. J. W. Cooley and J. W. Tukey, "An Algorithm for the Machine Calculation of Complex Fourier Series," Mathematics of Computation, Vol. 19, April 1965, pp. 297-301.
9. R. C. Singleton, "An Algorithm for Computing the Mixed Radix Fast Fourier Transform," IEEE Transactions on Audio and Electroacoustics, Vol. AU-17, No. 2, June 1969, pp. 93-103.
10. S. W. Golomb et al., Digital Communications, New Jersey: Prentice Hall, 1964, p. 53.
11. W. K. Pratt, J. Kane, and H. C. Andrews, "Hadamard Transform Image Coding," Proceedings of the IEEE, Vol. 57, January 1969, pp. 58-68.
12. J. E. Whelchel, Jr. and D. F. Guinn, "The Fast Fourier-Hadamard Transformation and Its Use in Signal Representation and Classifications," EASCON Record [1968 Electronics and Aerospace Systems Convention], pp. 561-571.
13. M. C. Pease, "An Adaptation of the Fast Fourier Transform for Parallel Processing," Journal of the Association for Computing Machinery, Vol. 15, No. 2, April 1968, pp. 252-264.
14. H. F. Harmuth, Transmission of Information by Orthogonal Functions, New York: Springer-Verlag, 1969.
15. J. L. Walsh, "A Closed Set of Orthogonal Functions," American Journal of Mathematics, Vol. 55, 1923, pp. 5-24.
16. G. S. Robinson, C. F. Chao, and R. L. Granger, "Speech Compression by Linear Transformations," COMSAT Laboratories Technical Memorandum CL-9-70, March 9, 1970.

17. A. Haar, "Zur Theorie der Orthogonalen Funktionensysteme," Mathematische Annalen, Vol. 69, 1910, pp. 331-371.
18. B. R. Gelbaum, "On the Functions of Haar," Annals of Mathematics, Vol. 51, No. 1, January 1950, pp. 26-36.
19. C. Watari, "A Generalization of Haar Functions," Tohoku Mathematical Journal, Vol. 8, 1956, pp. 286-290.
20. L. Collatz, Functional Analysis and Numerical Mathematics, New York: Academic Press, 1962.
21. H. C. Andrews and K. L. Caspari, "A Generalized Technique for Spectral Analysis," IEEE Transactions on Computers, Vol. C-19, No. 1, January 1970, pp. 16-25.
22. W. B. Davenport and W. L. Root, Random Signals and Noise, New York: McGraw-Hill, pp. 96-101, 1958.
23. A. Papoulis, Probability, Random Variables and Stochastic Processes, New York: McGraw-Hill, pp. 457-461, 1965.
24. T. Y. Young and W. H. Huggins, "On the Representation of Electrocardiograms," IEEE Transactions on Bio-Medical Electronics, July 1963.
25. S. Watanabe, "Karhunen-Loeve Expansion and Factor Analysis," Transactions of the Fourth Prague Conference on Information Theory, 1965, pp. 635-660.
26. H. C. Andrews and W. K. Pratt, "Transform Image Coding," Proceedings of the Symposium on Computer Processing in Communications, Polytechnic Institute of Brooklyn, April 8-10, 1969, pp. 63-83.

27. A. Habibi and P. Wintz, "Optimum Linear Transformations for Encoding 2-Dimensional Data," Purdue University, Technical Report No. TR-EE69-15, May 1969.
28. L. M. Goodman, "A Binary Linear Transformation for Redundancy Reduction," Proceedings of the IEEE (Letters), Vol. 55, No. 3, March 1967, pp. 467-468.
29. K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen-Loeve Expansion to Feature Selection and Ordering," IEEE Transactions on Computers, Vol. C-19, No. 4, April 1970, pp. 311-318.
30. K. Fukunaga and W. L. G. Koontz, "Representation of Random Processes Using the Finite Karhunen-Loève Expansion," Information and Control, Vol. 16, 1970, pp. 85-101.
31. J. Greenstadt, "The Determination of the Characteristic Roots of a Matrix by the Jacobi Method," Mathematical Methods for Digital Computers, A. Ralston and H. S. Wilf, ed., New York: John Wiley, pp. 84-91, 1960.
32. J. Ortega, "The Givens-Householder Method for Symmetric Matrices," Mathematical Methods for Digital Computers, A. Ralston and H. S. Wilf, ed., New York: John Wiley, pp. 94-115, 1967.
33. W. M. Goodall, "Television by Pulse Code Modulation," Bell System Technical Journal, Vol. 1, January 1951, pp. 33-49.
34. R. L. Carbrej, "Video Transmission over Telephone Cable Pairs by Pulse Code Modulation," Proceedings of the IRE, Vol. 48, September 1960, pp. 1546-1561.

35. F. W. Scoville and T. S. Huang, "The Subjective Effect of Spatial and Brightness Quantization in PCM Picture Transmission," Proceedings of the NEREM Conference Record, November 1965, pp. 234-235.
36. R. A. Bruce, "Optimum Pre-emphasis and De-emphasis Networks for Transmission of Television by PCM," IEEE Transactions on Communication Technology, Vol. COM-12, September 1964, pp. 91-96.
37. L. G. Roberts, "Picture Coding Using Pseudo-Random Noise," IRE Transactions on Information Theory, Vol. IT-8, February 1962, pp. 145-154.
38. P. F. Panter and W. Dite, "Quantization Distortion in Pulse Count Modulation with Nonuniform Spacing of Levels," Proceedings of the IRE, Vol. 39, January 1951, pp. 44-48.
39. J. Max, "Quantizing for Minimum Distortion," IRE Transactions on Information Theory, Vol. IT-16, March 1960, pp. 7-12.
40. H. A. Spang and P. M. Schultheiss, "Reduction of Quantizing Noise by Use of Feedback," IRE Transactions on Communications Systems, December 1962, pp. 373-380.
41. E. G. Kimme and F. F. Kuo, "Synthesis of Optimal Filters for a Feedback Quantization System," IEEE Transactions on Circuit Theory, Vol. CT-10, September 1963, pp. 405-413.
42. Kretzmer, E. R. "Statistics of Television Signals," Bell System Technical Journal, Vol. 31, July 1952, pp. 761-763.

43. L. E. Franks, "A Model for the Random Video Television Signals," Bell System Technical Journal, Vol. 45, April 1966, pp. 609-630.
44. W. K. Pratt and H. C. Andrews, "Transform Image Coding," USCEE Report No. 387, University of California, March 1970.
45. A. Habibi and P. Wintz, "Image Coding by Linear Transformation and Block Quantization," IEEE Transactions on Communications Technology, Vol. COM-19, No. 1, February 1971, pp. 50-62.
46. H. L. Van Trees, Detection, Estimation and Modulation Theory, Part I, New York: John Wiley, 1968.
47. W. D. Ray and R. M. Driver, "Further Decomposition of the Karhunen-Loève Series Representation of a Stationary Random Process," IEEE Transactions on Information Theory, Vol. IT-16, No. 6, November 1970, pp. 663-668.
48. W. R. Bennett, "Spectra of Quantized Signals," Bell System Technical Journal, Vol. 27, July 1968, pp. 446-472.
49. J. E. Abate, "Linear and Adaptive Delta Modulation," Proceedings of the IEEE, Vol. 55, No. 3, March 1967, pp. 298-308.